The ART of Divide and Conquer
An Innovative Approach to Improving the Efficiency of Adaptive Random Testing

Cliff Chow
The University of Hong Kong
Pokfulam, Hong Kong
cliffchow@hku.hk

Tsong Yueh Chen*
Swinburne University of Technology
Australia
tychen@swin.edu.au

T.H. Tse
The University of Hong Kong
Pokfulam, Hong Kong
thtse@cs.hku.hk

Abstract—Test case selection is a prime process in the engineering of test harnesses. In particular, test case diversity is an important concept. In order to achieve an even spread of test cases across the input domain, Adaptive Random Testing (ART) was proposed such that the history of previously executed test cases are taken into consideration when selecting the next test case. This was achieved through various means such as best candidate selection, exclusion, and diversity metrics. Empirical studies showed that ART algorithms make good use of the concept of even spreading and achieve 40 to 50% improvement in test effectiveness over random testing in revealing the first failure, which is close to the theoretical limit. However, the computational complexity of ART algorithms may be quadratic or higher, and hence efficiency is an issue when a large number of previously executed test cases are involved. This paper proposes an innovative divide-and-conquer approach to improve the efficiency of ART algorithms while maintaining their performance in effectiveness. Simulation studies have been conducted to gauge its efficiency against two most commonly used ART algorithms, namely, fixed size candidate set and restricted random testing. Initial experimental results show that the divide-and-conquer technique can provide much better efficiency while maintaining similar, or even better, effectiveness.

Keywords—adaptive random testing, divide and conquer, efficiency, effectiveness, software testing, test harness

I. INTRODUCTION

Test case selection is a major process in the development of test harnesses. Random testing [11] has been recognized as an important and useful method for test case selection. On the other hand, empirical studies have shown that failure-causing inputs, especially in numerical programs, tend to have contiguous failure regions [6]. Conceptually, if the current test cases cannot detect a failure, we should select the next test case to be far away from the test cases previously executed. Hence, evenly spreading the test cases across the input domain should provide a higher opportunity for revealing a failure.

Based on this intuition, Chen et al. [7] introduced the concept of Adaptive Random Testing (ART) to enhance the effectiveness of failure detection in random testing. The first ART method they proposed was the Fixed Size Candidate Set (FSCS) algorithm [6], [7], which generates a list of test cases as potential candidates, selects the best test case with the longest distance from a close neighborhood of previously executed test cases. Another proposed technique is the Restricted Random Testing (RRT), which selects the next test case as one that lies outside all the exclusive regions of previously executed test cases. Please refer to Section II for more details. Empirical studies show that ART algorithms, using the concept of even-spread, achieve 40 to 50% improvement in effectiveness over random testing (in terms of the expected number of test cases required to detect the first failure), which is close to the theoretical limit [8].

However, ART is less efficient than random testing because of the extra task of ensuring even spreading of test cases, where the efficiency is measured in terms of the time to generate a test case. This extra task involves the computation of history information from previous test cases because the determination of the i-th test case is related to the first to the (i−1)-th previously executed test cases. The computational complexity of ART algorithms is generally quadratic or higher. Hence, ART may not be more cost-effective than random testing when efficiency is taken into consideration.

In order to enhance the efficiency of existing ART algorithms, two techniques were proposed, namely, mirroring by Kuo [12] and forgetting by Chan et al. [2]. Conceptually speaking, both techniques can be applied to all ART methods. In the mirroring technique, the ART algorithm is applied to part of the input domain, and then the generated test cases are copied like mirror images to the remaining parts. In the forgetting technique, not all previously executed test cases are used.
to guide the generation of the next test case. The maximum number of previously executed test cases to be retained is specified by a memory parameter. On the other hand, empirical studies show that there is a trade-off between computational complexity and effectiveness [16]. Experimental results in the forgetting paper [2], for instance, show that for a given memory parameter, a similar performance in effectiveness cannot be maintained across different failure rates.

The motivation of the present paper is to design a new and innovative approach, known as divide-and-conquer, for reducing the cost of test case generation for ART algorithms, as well as preserving the performance in effectiveness.

The present study serves as an introduction to an innovative technique that improves on the efficiency of ART while preserving its effectiveness as far as possible. We will focus our experimental comparison with the two most commonly used ART algorithms, namely FSCS and RRT. Evaluations with respect to other ART algorithms will be left as future work.

The contribution of our proposed technique is fourfold: (a) It is an innovative enhancement of ART algorithms that greatly improves the computational complexity from quadratic or higher order to linear order. (b) It preserves the even spreading property of ART and largely maintains the performance in effectiveness. (c) It can be applied to all ART algorithms because it is an add-on technique. (d) It is independent of the dimension of the input domain.

The paper is organized as follows: Section II provides the background information on FSCS and RRT. Section III describes the proposed divide-and-conquer approach and explains the differences between the new technique and two existing efficiency improvement techniques. A comparison of the efficiency and effectiveness between selected ART algorithms and the same algorithms enhanced with the divide-and-conquer strategy is presented in Section IV. Finally, we present the conclusion in Section V.

II. BACKGROUND INFORMATION ON FSCS AND RRT

The Fixed Size Candidate Set (FSCS) algorithm for adaptive random testing, proposed by Chen et al. [6], [7], makes use of the simple intuition of selecting the best choice out of a fixed number of candidates to generate the next test case. Essentially, FSCS randomly generates $k$ candidates and then calculates the distance $dist_i$ between each candidate $c_i$ and its closest previously executed test case. The candidate $c_{best}$ with the longest distance, defined as $c_{best} = \{c_i | dist_i = max_{j \in \{1,2,...,k\}} dist_j\}$, represents the best candidate. It is selected as the next test case, and other candidates are discarded. This best candidate selection procedure is repeated until the first failure is revealed or the appropriate time limit has been reached. Since the computing operation calculates the distance between the candidates and all previously executed test cases, the computational complexity of FSCS is $\Theta(n^2)$ [6], [7].

Chan et al. [3] introduced Restricted Random Testing (RRT), which is based on the notion of exclusion. All exclusion zones are circular and equal in size. Let $A$ be the target exclusion zone and $n$ be the number of previously executed test cases. Each exclusion area is $A/n$, and the radius of each exclusion zone is $\sqrt{A/(\pi n)}$. For each round of test case generation, an exclusion zone has to be defined for every test case previously executed. The potential test cases are randomly generated until there is a potential test case outside all exclusion regions. It will be used as the next test case. The size of the exclusion zone will decrease when the number of previously executed test case increases. Mayer and Schneckenburger [15] estimated that the computational complexity of RRT is $\Theta(n^2 \log(n))$, assuming that the number of candidate test cases is logarithmic in the number of previously executed test cases. They validated their assumption empirically for the situation where the exclusion zones are circular, the total areas of the exclusion zones is 150% of the area of the input domain, and the number of test cases is no more than 500.

III. DIVIDE-AND-CONQUER APPROACH

A. Background

The proposed approach makes use of the standard concept of “divide and conquer” for breaking up a large problem into smaller sub-problems. It applies the exact ART algorithm to each of the sub-problems until the computational operation is too expensive for applying the ART algorithm again. It then further breaks up the sub-problems and repeats the procedure recursively until either a first failure is revealed or the time limit is reached.

While the target of the proposed divide-and-conquer approach is in alignment with two existing techniques for reducing the computational complexity of ART algorithms, namely, mirroring and forgetting, the concept of the new approach is different. In mirroring [12], the input domain is partitioned into disjointed sub-domains, and then the process selects a sub-domain as a source domain while the others are called mirror domains. The ART algorithm will only be applied to the source domain, and then the process uses a function to map the generated test cases from the source domain to all mirror domains. This mirroring procedure is repeated until a first failure is identified.

In forgetting [2], the researchers make use of the forgetting principle, which refers to “Human learning is often characterised by inaccurate retention or recall, termed forgetting” [2], to tackle the problem of the number of computational operations when the number of previously executed test cases grows. For a given memory parameter, the process forgets some previously executed test cases in order to improve the efficiency in generating the next test case.

The proposed divide-and-conquer approach uses bisectional division to break up the input domain into smaller sub-domains, and generates a next test case from each sub-domain. The divide-and-conquer approach does not forget any historical information from previously executed test cases. It only
ignores test cases outside the sub-domain under consideration in order to improve the efficiency of ART algorithms. This is a new and innovative approach for efficiency improvement in ART.

B. Proposed Approach

The approach starts with any ART algorithm and a specified integer known as the threshold, denoted by \( \lambda \). The determination of the threshold will be discussed later. Let \( D \) denote the input domain and \( T \) denote a set of previously executed test cases. After \( \lambda \) test cases have been generated in \( T \) from the input domain \( D \), the divide-and-conquer process is triggered. The process bisectionally divides each dimension of the input domain into equal-sized sub-domains \( \{D_1, D_2, \ldots, D_s\} \) such that \( \cup_{i=1}^{s} D_i = D \), where \( s = 2^{dn} \) denotes the number of sub-domains after each division process and \( dn \) denotes the dimension of the input space. The next step is to allocate test cases into their relevant sub-domains. This is achieved by putting all the sub-domains thus generated into a global queue \( \text{d_queue} \) and then assigning \( \lambda \) test cases to the relevant sub-domains. This is achieved by putting all the sub-domains thus generated into a global queue and selecting the least populated sub-domain \( D_i \) such that \( |T_i| = \min_{j \in \{1, 2, \ldots, s\}} |T_j| \), where \(||\) denotes the size of a set and \( T_i \) denotes the set of selected test cases in the sub-domain \( D_i \), that is, \( T_i = T \cap D_i \), from which the next test case will be generated.

Fig. 1. The operation of ART with divide-and-conquer.

As an example for illustration, consider a 2-dimension input domain \( D \). Fig. 1 shows the proposed divide-and-conquer approach in action. Suppose \( \lambda \) is 16. In Fig. 1(a), when the total number of test cases \(|T|\) in the input domain \( D \) reaches \( \lambda \), the divide-and-conquer process is triggered. In Fig. 1(b), input domains are bisectionally divided into four equal-sized sub-domains. After the test case reassignment process, the distribution of test cases are \(|T_1| = 4, |T_2| = 3, |T_3| = 4\), and \(|T_4| = 5\) in sub-domains \( D_1, D_2, D_3, \) and \( D_4 \) respectively. In accordance with the selection criterion, which chooses the least populated sub-domain to proceed, the next test case generation process is applied to sub-domain \( D_2 \). After generating one test case in \( D_2 \), the next test case should be generated from one of the three sub-domains \( \{D_1, D_2, D_3\} \), which have the same number of test cases previously executed (that is, 4). Then, the selection criterion is repeated until the numbers of test cases in all the sub-domains reach the threshold value \( \lambda \) again. In Fig. 1(c), the numbers of previously executed test cases in all sub-domains \( \{D_1, D_2, D_3, D_4\} \) reach the threshold value of 16, and the process starts the next bisectional division of all sub-domains. In Fig. 1(d), after the said bisectional division procedure, the process allocates all the sub-domains with a depth of 2 into a global queue (where depth denotes the depth of the bisectional division) and then repeats the selection criteria in order to find the least populated sub-domain \( D_{1.2} \) where \(|T_{1.2}|\) is 2

The algorithm for adaptive random testing with divide-and-conquer is as follows:

```
Algorithm 1. Adaptive Random Testing with Divide-and-Conquer (DC)

\[ \lambda \] = threshold for the divide-and-conquer process; /* The threshold is set by the human tester */
\[ D \] = the entire input domain; /* D */
\[ T \] = \{\}; /* a list to store the generated test cases in the input domain */
\[ depth \] = 0; /* The depth is set by the human tester */
\[ d\_queue \] = \{\}; /* a list to store the generated test cases in the input domain */
\[ t\_queue \] = \{\}; /* a list to store T */
\[ tmp\_d\_queue \] = \{\}; /* a temporary storage for D */
\[ tmp\_t\_queue \] = \{\}; /* a temporary storage for T */
while (stopping criteria not reached)
select the least populated sub-domain \( D_i \) from \( d\_queue \); /* that is, \( D_i \) such that \( |T_i| = \min_{j \in \{1, 2, \ldots, s\}} |T_j| \) */
\[ t = ART(D, T) \] /* generate a new test case by applying ART to D */
with \( T \) as the set of all test cases previously executed /* test the program using \( t \) as a test case;*/
if \( t \) reveals a failure
exit;
else
add \( t \) to \( T \);
end if;
if \( |T| = \lambda \)
remove \( D \) from \( d\_queue \);
remove \( T \) from \( t\_queue \);
start the bisectional division process on \( D \) and create \( s \) sub-domains \( D_{ij} \) such that \( \cup_{j=1}^{s} D_{ij} = D \), where \( s \) denotes the number of sub-domains after each division process;
create \( |T_i| \) lists to store the generated test cases in each sub-domain \( D_{ij} \);
end if;
for each \( t_i \) in \( T \) /* test case reassignment process */
if \( t_i \) is in \( D_{ij} \)
add \( t_i \) to \( T_{ij} \);
end if;
end for;
add all \( D_{ij} \) to \( tmp\_d\_queue \);
add all \( T_{ij} \) to \( tmp\_t\_queue \);
delete \( D \) and \( T \);
end if;
if \( t\_queue \) is empty
d\_queue = \( tmp\_d\_queue \);
t\_queue = \( tmp\_t\_queue \);
tmp\_d\_queue = \{\};
tmp\_t\_queue = \{\};
```
C. Determination of Threshold

The major concept of the proposed divide-and-conquer approach is to partition a large problem into smaller sub-problems when the number of previously executed test cases reaches the threshold \( \lambda \), in the sense that it tackles the smaller sub-domains individually rather than the entire input domain. If \( \lambda \) is set to be large, previously executed test cases should probably be evenly distributed, which is the property of ART, and there will be more or less the same number of test cases in each of the resultant sub-domains. However, the computational complexity will be high in this scenario. On the other hand, if \( \lambda \) is set to be too small, such as 4 in a 2-dimensional space, there will be a high chance that one of its 4 sub-domains may have no test case. This means that the next test case will be directly assigned to the sub-domain with no test case.

For the case of pure random testing, since the sub-domains are in equal-sized partitions, the determination problem can be viewed as a dice problem [9]. “How many dice must be rolled to have at least a 95% chance of rolling a one, a two, a three, a four, a five and a six?” The approach can be extended even further in accordance with the inclusion-exclusion principle [10], which is a mathematical counting technique to obtain the total number of elements in the union of two or more finite sets (such as \( |A \cup B \cup C| = |A| + |B| + |C| - |B \cap C| - |C \cap A| - |A \cap B| + |A \cap B \cap C| \), where \( A, B, \) and \( C \) are finite sets). The probability \( p(n) \) for having at least one test case residing in every resultant sub-domain in 2-dimensional space is:

\[
p(n) = \frac{3}{4} \sum_{j=0}^{3^n} (-1)^j \left( \frac{4}{4} \right)^j \left( 1 - \frac{1}{4} \right)^{n-j} \]

(1)

This probability exceeds 0.95 when \( n \geq 16 \). Table I shows some of the probabilities for various values of \( n \). In other words, there is 95% chance that at least one test case falls into every resultant sub-domain after the 16th test case generation in 2-dimensional input domain. Rewriting (1) in a general form in \( dn \)-dimensional input space, we have

\[
p(n) = \sum_{j=0}^{2^{dn}-1} (-1)^j \left( \frac{2^{dn}}{2^{dn}} \right)^j \left( 1 - \frac{1}{2^{dn}} \right)^{n-j} \]

(2)

Equation (2), however, is not applicable to adaptive random testing because ART does not produce independent and uniformly distributed random values. Having said that, (2) provides a rough guideline for determining the threshold for divide and conquer. More discussions and experimental results will be described in the next section.

IV. ANALYSIS AND EXPERIMENTAL RESULTS

A. Efficiency improvement

The determination of the threshold is an important decision. On one hand, the reduction of computational complexity will not be significantly improved if the threshold is set to be too high. On the other hand, if the threshold is assigned too small a value, the overall process will be dominated by sub-divisions of the input domain. Putting it to the extreme, suppose we set the threshold to be 1. The bisectonal division process will be executed whenever one test case is generated by random testing in the sub-domains. In other words, the ART algorithm is never applied in this scenario.

It is easy to see that the computational complexity of the divide-and-conquer approach contains two major components: The first component is the operation steps of the distance calculation for a specific ART algorithm in all input sub-domains, as shown in Figs. 1(a) and (c). The second is the process of reassigning all the test cases into newly created sub-domains by the bisectonal division procedure, as displayed in Figs. 1(b) and (d).

Consider the example in Fig. 1 that executes the FSCS-ART algorithm [6], [7] with divide-and-conquer technique in a 2-dimensional space such that the threshold \( \lambda \) is set to a constant value of 16. Referring to Section II, the complexity of distance computation in FSCS is \( \Theta(n^2) \). Since divide-and-conquer is triggered whenever that number of previously executed test cases has reached the threshold \( \lambda \), the asymptotic value of \( n \) will not exceed \( \lambda \), which is a constant. Hence, the complexity of distance computation in FSCS is also constant. Thus, after applying the divide-and-conquer approach, the computational complexity of ART algorithms becomes a linear order of the number of previously executed test cases.

On the other hand, another overhead is related to the process of dividing a large domain into smaller sub-domains and the assignment of test cases to each resultant sub-domain. Let \( m_{\text{first}} \) be the number of test cases to detect the first failure and \( s \) be the number of resultant sub-domains after dividing a domain. For instance, \( s = 4 \) in a 2-dimensional input space. The depth, which means the number of iterations of the division process, is given by \( m_{\text{first}} = 2^s \cdot 2^{\text{depth}-1} \). Rearranging the equation, we obtain:

\[
m_{\text{first}} = 2^s \cdot 2^{\text{depth}-1}
\]
\[ \text{depth} = \begin{cases} \frac{\log(m_{\text{first}}/\lambda)}{\log(s)} + 1 & m_{\text{first}} \geq \lambda \\ 0 & m_{\text{first}} < \lambda \end{cases} \] (3)

Considering the scenario in a 2-dimensional space, where \( s = 4 \). Let the threshold \( \lambda \) be 100. A depth of 0 will result when the number of test cases is in the range of \([0, 100)\), a depth of 1 will take place when the range is \([100, 400)\), a depth of 2 will result when the range is \([400, 1600]\), and so on. In general, since the depth can be found in (3), the overhead related to the bisectional division process and the assignment of test cases to each resultant sub-domain and the can be computed.

A comparison analysis was conducted on two commonly used ART algorithms, namely FSCS and RRT, with and without enhancement by the divide-and-conquer technique. In FSCS, the parameter for the number of potential candidates without enhancement by the divide-and-conquer technique. In FSCS-DC and RRT-DC compared with the basic FSCS and RRT, respectively.

The discrepancy metric indicates whether different regions in the input domain \( D \) have equal densities in test cases. The metric is formulated as:

\[ \text{discrepancy} = \max_{r \in \{1,2,...,1000\}} \left| \frac{T_r}{|T|} - \frac{D_r}{D} \right| \]

where \( D_r \) denotes a rectangular sub-domain of \( D \) whose size and location are randomly defined, and \( T_r \) denotes the set of test cases randomly selected from sub-domain \( D_r \), that is, \( T_r = T \cap D_r \). Thus, a low discrepancy value indicates that the test cases are equal in densities across all the sub-domains in \( D \).

The dispersion metric indicates whether there is a large empty region in the input domain \( D \), which is reflected by the maximum distance of any test case from its nearest neighbor. The metric is formulated as:

\[ \text{dispersion} = \max_{p \in \{1,2,...,|T|\}} \min_{q \in \{1,2,...,|T|\}} \text{dist}(t_p, t_q) \]

where the function \( \text{dist} \) denotes the Euclidean distance between two test cases. Thus, a low dispersion value indicates small empty regions in the input domain \( D \).

The empirical results in Figs. 2 and 3 demonstrate that significant efficiency improvements are achieved after applying the divide-and-conquer technique to either ART algorithm. The computational complexity is reduced. The reduction in computational complexity depends on the value of the threshold and the nature of the ART algorithm. Table II further shows that FSCS-DC and RRT-DC can perform better in terms of the discrepancy metric than the respective basic ART algorithms, FSCS and RRT. As expected, the dispersion diversity also indicates that there are larger empty regions for FSCS-DC and RRT-DC compared with the basic FSCS and RRT, respectively.

### Table II. Comparisons of Computational Complexity, Discrepancy, and Dispersion between FSCS and FSCS with DC and Between RRT and RRT with DC

<table>
<thead>
<tr>
<th>n</th>
<th>FSCS</th>
<th>FSCS-DC, ( \lambda = 100 )</th>
<th>FSCS-DC, ( \lambda = 50 )</th>
<th>FSCS-DC, ( \lambda = 10 )</th>
<th>RRT</th>
<th>RRT-DC, ( \lambda = 100 )</th>
<th>RRT-DC, ( \lambda = 50 )</th>
<th>RRT-DC, ( \lambda = 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>100</td>
<td>0.07</td>
<td>0.07</td>
<td>0.04</td>
<td>0.01</td>
<td>0.08</td>
<td>0.07</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>200</td>
<td>0.27</td>
<td>0.12</td>
<td>0.09</td>
<td>0.02</td>
<td>0.29</td>
<td>0.12</td>
<td>0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>400</td>
<td>1.04</td>
<td>0.33</td>
<td>0.15</td>
<td>0.05</td>
<td>1.23</td>
<td>0.33</td>
<td>0.15</td>
<td>0.05</td>
</tr>
<tr>
<td>800</td>
<td>4.10</td>
<td>0.54</td>
<td>0.36</td>
<td>0.10</td>
<td>5.01</td>
<td>0.55</td>
<td>0.37</td>
<td>0.10</td>
</tr>
<tr>
<td>1600</td>
<td>16.30</td>
<td>1.35</td>
<td>0.59</td>
<td>0.19</td>
<td>20.75</td>
<td>1.36</td>
<td>0.60</td>
<td>0.19</td>
</tr>
</tbody>
</table>

### B. Comparisons of Effectiveness Using F-measure

In this paper, the expected number of test cases to detect the first failure, usually referred to as the F-measure, is used as the effectiveness measure [4], [6]. F-measure is equal to \( 1/\theta \) or \(|D|/m\) for random testing with replacement, where \( m \) denotes the number of failures and \( \theta \) denotes the failure rate, such that \( m = |D|/\theta \). A lower value of the F-measure indicates a more effective testing approach. It is generally agreed that the F-measure is a more appropriate effectiveness measure for random testing and ART compared with other effectiveness metrics [1], [6].

There are three most typical failure patterns [4], [13], [14], namely, block, strip, and point patterns, as illustrated in Figs. 4a to 4c, respectively. In these figures, the bounding box indicates the input domain, and the colored block, strip, and dots represent the locations of failure-causing inputs.

The effectiveness for applying divide-and-conquer technique with respect to these typical failure patterns was studied via simulation. A 2-dimensional input domain was used, with randomly generated failure patterns. For the block pattern, a square was randomly chosen within the input domain such that its size yields the desired failure rate. For the strip pattern, two points on adjacent sides were picked, and a strip connecting the two points determined the failure region. For the point pattern, 50 non-overlapping circular regions were randomly chosen with equal radii.
### TABLE III. COMPARISONS OF F-MEASURES BETWEEN FSCS AND FSCS WITH DIVIDE-AND-CONQUER FOR BLOCK, STRIP, AND POINT PATTERNS

<table>
<thead>
<tr>
<th>Block Pattern</th>
<th>Failure rate</th>
<th>Expected F-measure of random testing (F_{RT})</th>
<th>Mean of FSCS (F_{FRT})</th>
<th>Mean of FSCS-DC (λ = 100) (F_{FRT-DC})</th>
<th>Mean of FSCS-DC (λ = 50) (F_{FRT-DC})</th>
<th>Mean of FSCS-DC (λ = 10) (F_{FRT-DC})</th>
<th>Mean of FSCS-DC (λ = 4) (F_{FRT-DC})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strip Pattern</td>
<td>0.01</td>
<td>100</td>
<td>92.1</td>
<td>93.6</td>
<td>94.1</td>
<td>95.7</td>
<td>95.7</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>200</td>
<td>188.2</td>
<td>190.1</td>
<td>189.8</td>
<td>192.3</td>
<td>200.6</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>500</td>
<td>475.3</td>
<td>486.0</td>
<td>484.7</td>
<td>488.7</td>
<td>527.2</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>1000</td>
<td>944.9</td>
<td>952.1</td>
<td>976.0</td>
<td>963.2</td>
<td>1070.5</td>
</tr>
<tr>
<td></td>
<td>0.0005</td>
<td>2000</td>
<td>1932.5</td>
<td>1922.4</td>
<td>1901.8</td>
<td>1928.6</td>
<td>2043.2</td>
</tr>
<tr>
<td>Point Pattern</td>
<td>0.01</td>
<td>100</td>
<td>100.6</td>
<td>99.7</td>
<td>100.4</td>
<td>100.9</td>
<td>108.3</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>200</td>
<td>200.3</td>
<td>199.9</td>
<td>201.6</td>
<td>201.8</td>
<td>218.8</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>500</td>
<td>496.8</td>
<td>495.9</td>
<td>494.7</td>
<td>501.2</td>
<td>525.8</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>1000</td>
<td>991.7</td>
<td>978.7</td>
<td>993.9</td>
<td>995.4</td>
<td>1070.5</td>
</tr>
<tr>
<td></td>
<td>0.0005</td>
<td>2000</td>
<td>1978.7</td>
<td>1965.7</td>
<td>1982.6</td>
<td>1973.9</td>
<td>2124.6</td>
</tr>
</tbody>
</table>

### TABLE IV. COMPARISONS OF F-MEASURES BETWEEN RRT AND RRT WITH DIVIDE-AND-CONQUER FOR BLOCK, STRIP, AND POINT PATTERNS

<table>
<thead>
<tr>
<th>Block Pattern</th>
<th>Failure rate</th>
<th>Expected F-measure of random testing (F_{RT})</th>
<th>Mean of RRT (F_{FRT})</th>
<th>Mean of RRT-DC (λ = 100) (F_{FRT-DC})</th>
<th>Mean of RRT-DC (λ = 50) (F_{FRT-DC})</th>
<th>Mean of RRT-DC (λ = 10) (F_{FRT-DC})</th>
<th>Mean of RRT-DC (λ = 4) (F_{FRT-DC})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strip Pattern</td>
<td>0.01</td>
<td>100</td>
<td>66.1</td>
<td>67.7</td>
<td>68.0</td>
<td>70.9</td>
<td>72.8</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>200</td>
<td>127.6</td>
<td>131.3</td>
<td>135.1</td>
<td>141.6</td>
<td>146.0</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>500</td>
<td>310.0</td>
<td>323.0</td>
<td>330.0</td>
<td>346.1</td>
<td>360.3</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>1000</td>
<td>614.9</td>
<td>645.4</td>
<td>659.1</td>
<td>698.5</td>
<td>727.2</td>
</tr>
<tr>
<td></td>
<td>0.0005</td>
<td>2000</td>
<td>1218.1</td>
<td>1289.2</td>
<td>1289.7</td>
<td>1365.2</td>
<td>1440.5</td>
</tr>
<tr>
<td>Point Pattern</td>
<td>0.01</td>
<td>100</td>
<td>91.4</td>
<td>92.9</td>
<td>93.6</td>
<td>94.1</td>
<td>95.0</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>200</td>
<td>184.6</td>
<td>190.7</td>
<td>189.9</td>
<td>189.8</td>
<td>192.2</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>500</td>
<td>479.1</td>
<td>479.3</td>
<td>482.4</td>
<td>479.8</td>
<td>484.3</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>1000</td>
<td>947.6</td>
<td>971.1</td>
<td>960.5</td>
<td>948.7</td>
<td>962.6</td>
</tr>
<tr>
<td></td>
<td>0.0005</td>
<td>2000</td>
<td>1906.8</td>
<td>1898.6</td>
<td>1939.0</td>
<td>1933.0</td>
<td>1933.0</td>
</tr>
</tbody>
</table>

### TABLE V. COMPARISONS OF F-MEASURES BETWEEN RANDOM TESTING (RT) AND RANDOM TESTING WITH DIVIDE-AND-CONQUER (RT-DC) FOR BLOCK, STRIP, POINT PATTERNS

<table>
<thead>
<tr>
<th>Block Pattern</th>
<th>Expected F-measure of RT (F_{FRT})</th>
<th>Mean of RT-DC (λ = 100) (F_{FRT-DC})</th>
<th>Mean of RT-DC (λ = 50) (F_{FRT-DC})</th>
<th>Mean of RT-DC (λ = 10) (F_{FRT-DC})</th>
<th>Mean of RT-DC (λ = 4) (F_{FRT-DC})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strip Pattern</td>
<td>0.01</td>
<td>100</td>
<td>98.5</td>
<td>98.0</td>
<td>90.0</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>200</td>
<td>195.7</td>
<td>192.3</td>
<td>179.8</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>500</td>
<td>491.1</td>
<td>489.5</td>
<td>449.9</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>1000</td>
<td>978.1</td>
<td>978.6</td>
<td>899.8</td>
</tr>
<tr>
<td></td>
<td>0.0005</td>
<td>2000</td>
<td>1942.2</td>
<td>1950.5</td>
<td>1789.5</td>
</tr>
<tr>
<td>Point Pattern</td>
<td>0.01</td>
<td>100</td>
<td>100.0</td>
<td>99.3</td>
<td>99.7</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>200</td>
<td>204.3</td>
<td>201.3</td>
<td>197.5</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>500</td>
<td>492.7</td>
<td>495.9</td>
<td>449.9</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>1000</td>
<td>1003.3</td>
<td>998.5</td>
<td>986.4</td>
</tr>
<tr>
<td></td>
<td>0.0005</td>
<td>2000</td>
<td>1976.8</td>
<td>1994.0</td>
<td>1965.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block Pattern</th>
<th>Failure rate</th>
<th>Expected F-measure of RT (F_{FRT})</th>
<th>Mean of RT-DC (λ = 100) (F_{FRT-DC})</th>
<th>Mean of RT-DC (λ = 50) (F_{FRT-DC})</th>
<th>Mean of RT-DC (λ = 10) (F_{FRT-DC})</th>
<th>Mean of RT-DC (λ = 4) (F_{FRT-DC})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strip Pattern</td>
<td>0.01</td>
<td>100</td>
<td>106.0</td>
<td>102.1</td>
<td>99.9</td>
<td>99.5</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>200</td>
<td>204.3</td>
<td>201.3</td>
<td>197.5</td>
<td>197.6</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>500</td>
<td>492.7</td>
<td>495.9</td>
<td>449.9</td>
<td>421.4</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>1000</td>
<td>1003.3</td>
<td>998.5</td>
<td>986.4</td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td>0.0005</td>
<td>2000</td>
<td>1976.8</td>
<td>1994.0</td>
<td>1965.2</td>
<td>1972.9</td>
</tr>
</tbody>
</table>
Fig. 2. Comparisons of computational complexity (in seconds) between FSCS and FSCS with divide-and-conquer with thresholds of 10, 50, and 100.

Fig. 3. Comparisons of computational complexity (in seconds) between RRT and RRT with divide-and-conquer with thresholds of 10, 50, and 100.

Fig. 4. Block, Strip, and Point failure patterns in 2-dimensional space.

The simulations were conducted with the following variables and values:

- Failure rates: 0.01, 0.005, 0.002, 0.001, and 0.0005
- Failure patterns: block, strip, and point
- Algorithms: FSCS, RRT, and random testing
- Thresholds: 100, 50, 10, and 4

For each combination of failure rate, failure pattern, algorithms, and threshold, 20,000 test runs were executed. The average F-measure for each exercise was collected. The same parameters for both the FSCS and RRT algorithms were used as per Section IV-A.

The experimental results in Tables III to IV demonstrate that the divide-and-conquer technique on top of ART performs similar, or even better, effectiveness than the original ART algorithms for the point failure pattern. The results also show that there is only a slight decline of 1–3% in effectiveness of the divide-and-conquer approach for the block and strip failure patterns.

Figs. 2 and 3 show that the F-measure slightly increases with efficiency improvement by reducing the threshold from 100 to 10. Tables III to IV show that the F-measure further increases if a small value of threshold such as 4 is assigned. These results imply that the effectiveness in FSCS-DC and RRT-DC becomes slightly poorer for smaller threshold values.

We also have studied the effect of divide-and-conquer technique to random testing. Table V illustrates that improvements in the F-measure for conducting random testing with divide-and-conquer can be significant. Future careful review will be required to verify whether divide-and-conquer can be used as a technique for improving the even spread of random test cases without the use of basic ART algorithms.

C. Threats to Validity

1) Internal validity.

This paper mainly serves to introduce an innovative efficiency improvement technique, namely divide-and-conquer. The experiments focus on simulations and may not represent a full variety of possible execution patterns of real-life faulty programs. However, in order to provide an “unbiased” cover, the experimental evaluations have been conducted on the three typical failure patterns, namely, block, strip, and point patterns, with wide ranges of failure rates. Various previous studies [3], [7], [16] have demonstrated that simulations of the three classic failure patterns can provide an objective validation of comparative merits.

A further study covering real-life faulty programs and including other ART algorithms will alleviate the threats to internal validity.

2) External validity.

There is one issue that may affect the external validity of the experiment. The Math::Random module, which is the Perl port of the C version of Randlib, is applied as the pseudorandom number generator to produce for the experiment a sequence of numbers that are approximately random. Randlib has been chosen because it comprises a large library of C routines for generating random numbers with uniform distribution under the UNIX and Linux environments. The use of other pseudorandom number generators may produce slightly different results.
3) Construct validity

There are two commonly used metrics to evaluate the effectiveness performance of testing algorithms, namely F-measure and P-measure, which refer to the expected number of test cases to detect the first failure and the probability of detecting at least one failure, respectively. To serve the purpose of introducing an innovative technique in this paper, F-measure is more appropriate as an effectiveness performance metric in order to synchronize and compare with various previous studies [3], [7], [15]. Furthermore, empirical studies [1], [16] have consistently shown that ART outperforms RT with respect to the P-measure. For these reasons, F-measure rather than P-measure has been used for the effectiveness performance metric in the experimental evaluation.

V. CONCLUSION

Previous empirical studies demonstrated that ART makes better use of the even-spreading concept than random testing and needs fewer test cases to reveal the first failure. In contrast, the computation cost for generating test cases in ART has been reported to be high. We propose a divide-and-conquer technique to address this problem. We make use of the intuition of breaking up a large problem into smaller subproblems and specify a threshold to limit the computational growth when a large number of previously executed test cases are involved in an ART algorithm. On the other hand, although the use of the threshold hides some of the history information from previously executed test cases, the subdomains after the bisectional division process can be treated as temporary space to store the past history and further build up an even spreading of test cases. In addition, the divide-and-conquer technique has an equal-number-equal-size characteristic to ensure that each sub-domain is equal in size and contains the same number of test cases, so that every sub-domain has the same probability in revealing the first failure. Consequently, the divide-and-conquer technique can significantly reduce the computational complexity of common ART algorithms such as FSCS and RRT while largely maintaining the effectiveness.

For future work, real-life programs will be utilized to enrich the evaluation. A more comprehensive comparison analysis between divide-and-conquer and other efficiency improvement techniques will also be conducted.

ACKNOWLEDGMENT

This work is supported in part by a linkage grant of the Australian Research Council (project no. LP100200208) and grants of the General Research Fund of the Research Grants Council of Hong Kong (project nos. 717811 and 716612).

REFERENCES