

Identifying minimal failure-inducing schemas for multiple faults^{*}

[Extended Abstract][†]

Xintao Niu Changhai Nie
State Key Laboratory for Novel Software Technology State Key Laboratory for Novel Software Technology
Nanjing University Nanjing University
China, 210093 China, 210093
niuxintao@smail.nju.edu.cn changhainie@nju.edu.cn

ABSTRACT

Combinatorial testing (CT) is proven to be effective to reveal the potential failures caused by the interaction of the inputs or options of the system under test (SUT). A key problem in CT is to isolate the failure-inducing interactions in SUT as it can facilitate the debugging effort by reducing the scope of code that needs to be inspected. Many algorithms have been proposed to identify the failure-inducing interactions in SUT, however, most of these studies either just consider the condition of one fault in SUT or ignore masking effects among multiple faults which can bias their identified results. In this paper, we analysed how the masking effect of multiple faults impact on the isolation of failure-inducing interactions. We further give a strategy of selecting test cases to alleviate this impact. The key to the strategy is to prune these test cases that may trigger masking effect and generate no-masking-effect ones to test the interactions supposed to be tested in these pruned test cases. The test-case selecting process repeated until we get enough information to isolate the failure-inducing interactions in SUT. We conducted some empirical studies on several open-source GNU software. The result of the studies shows that multiple faults do exist in real software and our approach can assist combinatorial-based failure-inducing identifying methods to get a better result when handling multiple faults in SUT.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
D.2.8 [Software Engineering]: Metrics—*complexity mea-*

^{*}(Does NOT produce the permission block, copyright information nor page numbering). For use with ACM_PROC_ARTICLE-SP.CLS. Supported by ACM.

[†]A full version of this paper is available as *Author's Guide to Preparing ACM SIG Proceedings Using L^AT_EX₂ ϵ and BibT_EX* at www.acm.org/eaddress.htm

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WOODSTOCK '97 El Paso, Texas USA

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

asures, performance measures

General Terms

Theory

Keywords

Minimal failure-inducing schemas, Masking effect

1. INTRODUCTION

With the increasing complexity and size of modern software, many factors, such as input parameters and configuration options, can influence the behaviour of the system under test (SUT). The unexpected faults caused by the interaction among these factors can make testing such software a big challenge if the interaction space is large. Combinatorial testing can systematically sample the interaction space and select a relatively small size of test cases that cover all the valid iterations with the number of factors involved in the interaction no more than t .

Once failures are detected by some test cases, CT then needs to isolate the failure-inducing interactions. This is important as it can facilitate the debugging effort by reducing the code scope that needed to be inspected. Many algorithms have been proposed to identify the failure-inducing interactions in SUT, which ranges from using classification tree model, generating additional test cases, ranking suspicious interactions based on rules, using graphic-based deduction and so on. They can be partitioned into two categories: for single fault and for multiple faults. The former aims at identifying the failure-inducing interactions in SUT with only one fault, i.e., the oracle of the test cases is either pass or fail, while the latter can distinguish the fail oracles into multiple kinds according to the fault information.

Put aside the performance these approaches exhibited for single fault, however, in our recent studies on several GNU open-source software, we find these methods didn't behave as expected when these software contain multiple faults. The main reason why these methods fail to behave normally is that they didn't consider the masking effect that may happen among different faults. Take the Linux command-Grep for example, we noticed that there are two different faults reported in the bug tracker system. The first one claims that Grep incorrectly matches unicode patterns with ' $\backslash<\backslash>$ ', while the second one claims an incompatibility between op-

tion 'c' and 'o'. When we put this two scenario into one test case only one fault information will be observed, which means another fault is masked by the observed one. Obviously this fact do trouble these algorithms as it make them unable to judge whether the test case under testing trigger only the fault observed or triggered both two faults. As a result, it will make wrong analysis in the failure-inducing interactions for these masked faults.

One insight in this paper is that we cannot completely get away from the masking effect even if we do exhaustive testing. Further more, both ignoring the masking effect and regarding multiple faults as one fault is harmful for our identifying process. One remedy to alleviate this problem is to select test cases to reduce the bias, i.e., select test cases that without suffering the masking effect to get a partial but masking-avoiding result. However, most selecting strategy in CT is to cover the interactions with the number of test cases as small as possible. So in this paper we proposed a strategy for selecting test cases with the aim to alleviate the masking effect.

The key to the strategy is to prune test cases that may trigger a masking effect and then select or generate other test cases to test the interactions which were supposed to be tested in these pruned ones. We will keep these pruned test cases for the next iteration to isolate the failure-inducing interactions in these test cases. The process repeated until we characterize all the interactions for each fault. A point need to be noted is that these interactions supposed to be tested in the pruned test cases for one interaction vary in different algorithms we adopted. For example, for the classified tree method, we will generate more test cases that keep the same coverage, and for the one fact one time, we will generate test cases to keep the same.

We tracked several open-source software in the bug tracker system and find that multiple faults do exist in software of the same version. To evaluate the effectiveness of our approach, we take software benchmark from SIR with the ability to inject bugs into the source code. Based on this controllable subject we choose three different failure-inducing interactions identifying algorithms to compare the performance between the original algorithm and its variation extended with our strategy. The results shows that the algorithm extend with our test cases selecting strategy can perform better than before.

The main contributions of this paper are:

1. We studied the impact on the isolation of the failure-inducing interactions when the SUT contain multiple faults which can mask each other.
2. We proposed a strategy of selecting test cases to reduce the impact of masking effect.
3. We empirically studies our strategy and find that our approach can get a better result.

Rest of paper is organised as follows: section 2 gives a simple example to motivate our work. Section 3 give some background of the work. Section 4 describe our approach in detail. Section 5 illustrate the experiment and reports the result. Section 6 discusses the related works. Section 7 provides some concluding remarks.

2. MOTIVATION EXAMPLE

Following we have construct a example to illustrate the motivation of our approach. Assume we have a method *foo* which has four input parameters : *a*, *b*, *c*, *d*. The types of these four parameters are all integers and the values that they can take are: $d_a = \{7, 11\}$, $d_b = \{2, 4, 5\}$, $d_c = \{4, 6\}$, $d_d = \{3, 5\}$ respectively. The detail code of this method is listed as following:

```
public static float foo(int a, int b, int c, int d){
    //step 1 will cause a exception when b == c
    float x = (float)a / (b - c);

    //step 2 will cause a exception when c < d
    float y = Math.sqrt(c - d);

    return x+y;
}
```

Inspecting the simple code above, we can find two faults: First, in the step 1 we can get a *ArithmeticException* when *b* is equal to *c*, i.e., *b* = 4 & *c* = 4, that makes division by zero. Second, another *ArithmeticException* will be triggered in step 2 when *c* < *d*, i.e., *c* = 4 & *d* = 5, which makes square roots of negative numbers. So the expected MFSs in this example should be (-, 4, 4, -) and (-, -, 4, 5).

Traditional MFS identifying algorithms do not consider the detail of the code. They take black-box testing of this program, i.e., feed inputs to those programs and execute them to observe the result. The basic justification behind those approaches is that the failure-inducing schema for a particular fault must only appear in those inputs that trigger this fault. As traditional MFS identifying algorithms aim at using as small number of inputs as possible to get the same or approximate result as exhaustive testing, so the results derive from a exhaustive testing set must be the best that these MFS identifying approaches can reach. Next we will illustrate how exhaustive testing works on identifying the MFS in the program.

We first generate every possible inputs as listed in the Column "test inputs" of table 1, and execute them to get the result listed in Column "result" of table 1. In this Column, "PASS" means that the program runs without any exception under the inputs in the same row. "Ex 1" indicate that the program encounter a exception corresponding to the step 1 and "Ex 2" indicate the program trigger a exception corresponding to the step 2. According to data listed in table 1, we can deduce that that (-, 4, 4, -) must be the MFS of Ex 1 as all the inputs triggered Ex 1 contain this schema. Similarly, the schema (-, 2, 4, 5) and (-, 3, 4, 5) must be the MFSs of the Ex 2. We listed the MFSs and its corresponding exception in table 2.

Note that we didn't get the expected result with traditional MFS identifying approaches for this case. The MFSs we get for Ex 2 are (-,2,4,5) and (-,3,4,5) respectively instead of the expected schema (-,-,4,5). So why we can't identify the MFS (-,-,4,5)? The reason lies in the two inputs: input 6. (7,4,4,5) and input 18. (11,4,4,5). This two inputs contain the schema (-,-,4,5), but didn't trigger the Ex 1, instead, the Ex 2 was triggered.

Now let us get back to the source code of *foo*, we can find that if Ex 1 are triggered, it will stop executing the remaining code and report the exception information. In another word, Ex 1 have a higher level than Ex 2 so that Ex

Table 1: test inputs and their corresponding result

id	test inputs	result
1	(7, 2, 4, 3)	PASS
2	(7, 2, 4, 5)	Ex 2
3	(7, 2, 6, 3)	PASS
4	(7, 2, 6, 5)	PASS
5	(7, 4, 4, 3)	Ex 1
6	(7, 4, 4, 5)	Ex 1
7	(7, 4, 6, 3)	PASS
8	(7, 4, 6, 5)	PASS
9	(7, 5, 4, 3)	PASS
10	(7, 5, 4, 5)	Ex 2
11	(7, 5, 6, 3)	PASS
12	(7, 5, 6, 5)	PASS
13	(11, 2, 4, 3)	PASS
14	(11, 2, 4, 5)	Ex 2
15	(11, 2, 6, 3)	PASS
16	(11, 2, 6, 5)	PASS
17	(11, 4, 4, 3)	Ex 1
18	(11, 4, 4, 5)	Ex 1
19	(11, 4, 6, 3)	PASS
20	(11, 4, 6, 5)	PASS
21	(11, 5, 4, 3)	PASS
22	(11, 5, 4, 5)	Ex 2
23	(11, 5, 6, 3)	PASS
24	(11, 5, 6, 5)	PASS

Table 2: Identified MFSs and their corresponding Exception

MFS	Exception
(-, 4, 4, -)	Ex 1
(-, 2, 4, 5)	Ex 2
(-, 3, 4, 5)	Ex 2

Table 3: expected MFSs and their corresponding Exception

MFS	Exception
(-, 4, 4, -)	Ex 2
(-, -, 4, 5)	Ex 1

1 may mask Ex 2. With this information, we can suppose that for the input (7,4,4,5) and (11,4,4,5), Ex 1 may masked Ex 2. Then we exam the schema $(-, -, 4, 5)$, we can find all the test inputs contain $(-, -, 4, 5)$ will trigger exception 1, except these test cases trigger Ex 2 first. So we can conclude that $(-, -, 4, 5)$ should be the causing schema of the Exception 1. So the MFS information will be updated to table 3 which is identical to the expected result.

So in this paper, we need to analysis the priority among the faults in the SUT and use the information to assist the MFS identifying algorithms to make the result more accurate and clearer.

3. PRELIMINARY

Before we talk about our approach, we will give some formal definitions and background first, which is helpful to understand the description of our approach.

3.1 Combinatorial testing

Assume that the SUT (software under test) is influenced by n parameters, and each parameter c_i has a_i discrete values from the finite set V_i , i.e., $a_i = |V_i|$ ($i = 1, 2, \dots, n$). Some of the definitions below are originally defined in .

Definition 1. A test configuration of the SUT is an array of n values, one for each parameter of the SUT, which is denoted as a n -tuple (v_1, v_2, \dots, v_n) , where $v_1 \in V_1, v_2 \in V_2 \dots v_n \in V_n$.

Definition 2. We consider the fact that abnormal executing of the SUT as a *fault*. It can be a exception, a compilation error, a mismatched assertion or a constraint violation.

Definition 3. The *priority* is a function indicate the priority relationship between two faults. Specifically, we take $Priority(F_a, F_b) = 1$ as that fault F_a has a higher level than F_b , which means that if F_a were triggered, it will omit the code that may trigger F_b . And $Priority(F_a, F_b) = -1$ indicate that F_a has a lower level than F_b . Finally, we take $Priority(F_a, F_b) = 0$ as that there is no priority relationship between F_a and F_b , in another word, neither F_a will mask F_b nor F_b will mask F_a .

Definition 4. For the SUT, the n -tuple $(-, v_{n_1}, \dots, v_{n_k}, \dots)$ is called a k -value *schema* ($k > 0$) when some k parameters have fixed values and the others can take on their respective allowable values, represented as "-". In effect a test configuration its self is a k -value *schema*, which k is equal to n . Furthermore, if a test configuration contain a *schema*, i.e., every fixed value in this schema is also in this test configuration, we say this configuration hit this *schema*.

Definition 5. let s_l be a l -value schema, s_m be an m -value schema for the SUT and $l \leq m$. If all the fixed parameter values in s_l are also in s_m , then s_m *subsumes* s_l . In this case we can also say that s_l is a *sub-schema* of s_m and s_m is a *parent-schema* of s_l .

Definition 6. If all test configurations except these configurations triggered a higher level fault contain a schema, say S_a , trigger a particular fault, say F_a , then we call this schema S_a the *faulty schema* for F_a . Additionally, if none sub-schemas of S_a is the *faulty schema* for F_a , we will call the schema S_a the *minimal faulty schema* for F_a (MFS for short).

Note that, traditional MFS definition didn't consider the priority relationship among faults, so these definition will not take the schema as a MFS for some particular fault if some test configuration contain this schema doesn't trigger this fault.

The target of traditional MFS identifying algorithms is to find the MFSs for the faults of a SUT. For by doing that can help the developers reduce the scope of source code that needed to inspect to debug the software. Note that our discuss is based on the SUT is a deterministic software, i.e., SUT execute under a test configuration will not pass one time and fail another time. The non-deterministic problem will complex our test scenario, which, however is beyond the scope of this paper.

3.2 Dominate tree

fault. level. given this definition, we can also take the constraint as a fault, usually it will have the highest level, for that if we take a combination which is constraint, it will may didn't compile at all, so that we can't exam any code in this software.

test configuration?

schema. faulty, healthy. we need to identify the minimal faulty schema, that is MFS.

note that in our paper, we just consider the failure that is deterministic.

4. THEORY FOUNDATION

For a SUT, assume it has n different faults: $F_i (1 \leq i \leq L)$, they can be distinguished by the fault information come with them (such as exception traces or fatal code). Without loss of generality, we assume monotonicity of the level of faults ($Level(F_1) \geq Level(F_2) \geq \dots \geq Level(F_L)$).

Let S_t denote all the schemas in the test configuration t . For example, If $T = (1, 1, 1)$. Then S_T is $\{(1, 1, 1), (1, 1, -), (1, -, 1), (-, 1, 1), (-, -, 1), (-, 1, -)\}$.

Similarly, Let S_T denote all the schemas in a suite of test configurations $-T$. In fact, $S_T = \bigcup_{i=1}^n S_{t_i}$.

The notation T_{F_i} denote a suite of test configurations that each test configuration in it will trigger fault F_i . Additionally, let T_P denote the suite of test configurations that passed the testing without triggering any faults, and let T_F indicate the suite of test configurations that each one in this set will trigger some type of fault. It is obviously that $T_F = \bigcup_{i=1}^L T_{F_i}$ and $T_F \cup T_P = T_{all}$, In which, T_{all} denote all the possible test configurations that can generate to test the SUT.

Then let $M_{F_i} (1 \leq i \leq L)$ denote the set of MFS for F_i . It is noted that for F_i there may be more than one MFS, so we use set of MFS instead one MFS for a fault.

As MFS is also a schema, so we can easily find the followed properties: $M_{F_i} \in S_{T_{F_i}}$

4.1 Regard as one fault

Table 4: a simple example

id	test inputs	result
1	(0, 0, 0)	Err 1
2	(0, 0, 1)	PASS
3	(0, 1, 0)	Err 1
4	(0, 1, 1)	PASS
5	(1, 0, 0)	Err 1
6	(1, 0, 1)	PASS
7	(1, 1, 0)	Err 2
8	(1, 1, 1)	Err 2

Table 5: previous work

id	test inputs	result
1	(0, 0, 0)	FAIL
2	(0, 0, 1)	PASS
3	(0, 1, 0)	FAIL
4	(0, 1, 1)	PASS
5	(1, 0, 0)	FAIL
6	(1, 0, 1)	PASS
7	(1, 1, 0)	FAIL
8	(1, 1, 1)	FAIL

In our previous work (TOSEM), we didn't distinguish the types of fault, i.e., we treat all the faults as one failure. In this circumstance, our MFS identifying process can be formulated as:

$$M_F = \{s_i | s_i \in S_{T_F} - S_{T_P} \text{ and } \nexists s_j \prec s_i \text{ s.t. } s_j \in S_{T_F} - S_{T_P}\}$$

With this formula we cannot find the MFS for a particular fault, further more, as it force to merge all the MFS of, it may get a wrongly result. For example, for SUT(3,2), Assume (0,-,0) and (-,0,0) are the MFS of err 1 and (1,1,-) and (1,-,0) is the MFS of err 2. Err 1's level is higher than Err 2. The test configurations and result is listed in table 4.

Then using our previous work, firstly, we will treat the test result as listed in 5.

And then we will wrongly identified as (1,1,-) and (-,0,0) for the fault.

4.2 Do not consider the masking effect

To solve this problem, a naturally solution is to distinguish these faults, and for a particular fault, say F_i , we treat the test configurations trigger this fault as failure test configuration and the remained test configurations (consist of pass and other faults test configurations) will be regarded as pass.

So for this idea, the identifying process for a particular fault, say, F_i can be formulated as follows:

Let

$$S_{ca} = S_{T_{F_i}} - S_{T_P} - \bigcup_{j=1 \& j \neq i}^L S_{T_{F_j}}$$

$$M_F = \{s_i | s_i \in S_{ca} \text{ and } \nexists s_j \prec s_i \text{ s.t. } s_j \in S_{ca}\}$$

Using this we will treat the table 4 as the following table 6, then we will get the result as : (0,-,0) and (-,0,0) are the MFS of err 1 and (1,1,-) is the MFS of err 2. It is very approximate to the solution except that it loose the (1,-,0) for err 2. This is because in fact when we look at the 5th test

Table 6: do not consider the masking effect

ERR 1			ERR 2		
id	test inputs	result	id	test inputs	result
1	(0, 0, 0)	FAIL	1	(0, 0, 0)	PASS
2	(0, 0, 1)	PASS	2	(0, 0, 1)	PASS
3	(0, 1, 0)	FAIL	3	(0, 1, 0)	PASS
4	(0, 1, 1)	PASS	4	(0, 1, 1)	PASS
5	(1, 0, 0)	FAIL	5	(1, 0, 0)	PASS
6	(1, 0, 1)	PASS	6	(1, 0, 1)	PASS
7	(1, 1, 0)	PASS	7	(1, 1, 0)	FAIL
8	(1, 1, 1)	PASS	8	(1, 1, 1)	FAIL

configuration (1,0,0), it has already triggered the err 1 which have a higher level than err 2 so that err 2 is not triggered. But this approach just regard the 5th test configuration as a passed test configuration as it does not trigger err 2.

4.3 Test configurations with higher level fault masked every lower level fault

So what if we just consider the configurations triggered higher fault as having triggered the lower fault, in other words, we defaulted think that for each test configurations triggering a fault, say, F_i , it has masked all the faults has a level lower than it, i.e., these faults $F_j, i < j \leq L$ will be triggered in these configurations if the fault F_i does not trigger. For this idea, the identifying process for a particular fault with considering masking effect, say, F_i can be formulated as follows:

Let

$$S_{ca} = \bigcup_{j=1}^i S_{T_{F_j}} - S_{T_P} - \bigcup_{j=i+1}^L S_{T_{F_j}}$$

$$M_F = \{s_i | s_i \in S_{ca} \text{ and } \nexists s_j \prec s_i \text{ s.t. } s_j \in S_{ca}\}$$

We can see the different part with the previous formula is that it just minus the schemas in these passing test configurations and these test configurations triggering a lower level than F_i . With this, we will treat the table 4 as : table 7. This time we get (0,-,0) and are the MFS of err 1, (1,1,-) and (-,-,0) for the err 2. Obviously it is still not the correct answer.

This is because we look at 1th, 3th test configuration, actually they have trigger the err 1 which has the higher level than err 2, but it does not mask the err 2 as this two test configurations didn't contain the MFS for err 2 we set at first. So they should be regard as pass test configuration for err 2, but this approach didn't know this, they just set all the test configurations that trigger higher level fault as the fail test configuration for the under test fault.

4.4 Ideal solution

To get an ideal solution, we should make the followed assumption:

For a particular fault, say, F_i , with these test configurations triggered higher fault are $\bigcup_{j=1}^{i-1} T_{F_j}$. Assume we have known previously that in this set some test configurations will trigger F_i if the higher fault will not triggered, we label this set as

T_{tri-F_i} (this part is needed because of if there are more than two faults in a SUT, the higher fault may even make the lower fault don't appear)

Table 7: over mask example

ERR 1			ERR 2		
id	test inputs	result	id	test inputs	result
1	(0, 0, 0)	FAIL	1	(0, 0, 0)	FAIL
2	(0, 0, 1)	PASS	2	(0, 0, 1)	PASS
3	(0, 1, 0)	FAIL	3	(0, 1, 0)	FAIL
4	(0, 1, 1)	PASS	4	(0, 1, 1)	PASS
5	(1, 0, 0)	FAIL	5	(1, 0, 0)	FAIL
6	(1, 0, 1)	PASS	6	(1, 0, 1)	PASS
7	(1, 1, 0)	PASS	7	(1, 1, 0)	FAIL
8	(1, 1, 1)	PASS	8	(1, 1, 1)	FAIL

Table 8: ideal solution

ERR 1			ERR 2		
id	test inputs	result	id	test inputs	result
1	(0, 0, 0)	FAIL	1	(0, 0, 0)	PASS
2	(0, 0, 1)	PASS	2	(0, 0, 1)	PASS
3	(0, 1, 0)	FAIL	3	(0, 1, 0)	PASS
4	(0, 1, 1)	PASS	4	(0, 1, 1)	PASS
5	(1, 0, 0)	FAIL	5	(1, 0, 0)	FAIL
6	(1, 0, 1)	PASS	6	(1, 0, 1)	PASS
7	(1, 1, 0)	PASS	7	(1, 1, 0)	FAIL
8	(1, 1, 1)	PASS	8	(1, 1, 1)	FAIL

And some test configurations in this set will not trigger F_i . We label them as:

$$T_{-tri-F_i}$$

$$\text{Obviously, } T_{tri-F_i} \cup T_{-tri-F_i} = \bigcup_{j=1}^{i-1} T_{F_j}.$$

At last, we should do as the following to get the ideal computing formula:

Let

$$S_{ca} = S_{T_{tri-F_i}} + S_{T_{F_i}} - S_{T_P} - S_{T_{-tri-F_i}} - \bigcup_{j=i+1}^L S_{T_{F_j}}$$

$$M_F = \{s_i | s_i \in S_{ca} \text{ and } \nexists s_j \prec s_i \text{ s.t. } s_j \in S_{ca}\}$$

So still for the example table 4, we will first get $T_{tri-F_i} = \{(1,0,0)\}$ and $T_{tri-F_i} = (0, 0, 0), (0, 1, 0)$

And then we will treat table 4 as table 8, so this time we will accurately get the expected result:

(0,-,0) and (-,0,0) are the MFS of err 1 and (1,1,-) and (1,-,0) is the MFS of err 2.

4.5 limitations in practice

In practice, we cannot use the ideal formula to compute the MFS for each fault for there are three main limitations which we will encounter:

1. We cannot execute all the test configurations if the parameters and their values it too much for the SUT, that is in practice, $T_F \cup T_P \neq T_{all}$
2. If we find the faults, we can't judge the levels of these faults just using black-box testing method.
3. Even if we have known the levels of each faults, we still cannot decide the value of $S_{T_{tri-F_i}}$ and $S_{T_{-tri-F_i}}$ without fixing the higher level fault than F_i and re-executing the test configurations.

5. APPROACH DESCRIPTION

As we cannot get an ideal solution because of the three limitations proposed in the previous section, our target then is to find a practical solution to improve the efficiency for the existing MFS identifying algorithms when facing multiple faults, i.e., lowering variance of identifying MFS in a SUT with multiple faults as much as possible.

To get the target, first let's get back to the traditional MFS identifying algorithms to see how they works. In fact, they all can be represent as the following formula:

Let

$$S_{ca} = S_{T'_{F_i}} - S_{T'_P}$$

$$M_F = \{s_i | s_i \in S_{ca} \text{ and } \nexists s_j \prec s_i \text{ s.t. } s_j \in S_{ca}\}$$

A noted point is that $T'_{F_i} \subseteq T_{F_i}$ and $T'_P \subseteq T_P$ as that we can't execute all the possible test configurations in a SUT in practice when the scale of the configuration space is big.

The difference among these algorithms are just at T'_{F_i} and T'_P . However, what the difference in detail is not the point in this paper. We should also note that when the SUT just have one fault, these algorithms all can get a good result. So what we want to do is to improve these algorithms when the SUT can have multiple faults.

As we have mentioned in the previous section, the $S_{T'_{F_i}} - S_{T'_P}$ is not completed. This is because there may be some failure-inducing schemas in some T_{F_j} we did not add and there may be some healthy schemas in T_{F_j} we did not minus. This two factors we can't improve, however, because we neither know the levels of these faults nor know value of $S_{T_{tri-F_i}}$ and $S_{T_{-tri-F_i}}$. So what we can do is just to increase the number of T'_{F_i} and T'_P to increase the accurately of identifying the MFS of a particular fault F_i .

5.1 Replace test configuration that trigger unexpected fault

The basic idea is to discard the test configurations that trigger other faults and generate other test configurations to represent them. These regenerate test configurations should either pass the executing or trigger F_i . The replacement must fulfil some criteria, such as for CTA, the input for this algorithm is a covering array, and if we replace some test configuration in it, we should ensure that the covering rate is not changed.

Commonly, when we replace the test configuration that trigger unexpected fault with a new test configuration, we should keep some part in the original test configuration, we call this part as *fixed part*, and mutant other part with different values from the original one. For example, if a test configuration (1,1,1,1) triggered Err 2, which is not the expected Err 1, and the fixed part is (-,-,1,1), then we may regenerate a test configuration (0,0,1,1) which will pass or trigger Err 1.

The *fixed part* can be the schemas that only appear in the original test configuration which other test configurations in the covering array did not contain(for the algorithm take covering array as the input: CTA), or the factors that should not be changed in the OFOT algorithms, or the part that should not be mutant of the test configuration in the last iteration(FIC_BS).

The process of replace a test configuration with a new one with keeping some fixed part is depicted in Algorithm 1:

Algorithm 1 replace test configurations that trigger unexpected fault

Input: $t_{original}$ \triangleright original test configurations
 F_i \triangleright fault type
 s_{fixed} \triangleright fixed part
 $Param$ \triangleright values that each option can take

Output: t_{new} \triangleright the regenerate test configuration

```

1: while not MeetEndCriteria() do
2:    $s_{mutant} \leftarrow t_{original} - s_{fixed}$ 
3:   for each  $opt \in s_{mutant}$  do
4:      $i = getIndex(Param, opt)$ 
5:      $opt \leftarrow opt' \text{ s.t. } o \in Param[i] \text{ and } opt' \neq opt$ 
6:   end for
7:    $t_{new} \leftarrow s_{fixed} \cup s_{mutant}$ 
8:    $result \leftarrow execute(t_{new})$ 
9:   if  $result == PASS$  or  $result == F_i$  then
10:    return  $t_{new}$ 
11:   else
12:     continue
13:   end if
14: end while
15: return null

```

The inputs for this algorithm consists of a test configuration which trigger an unexpected fault - $t_{original}$, the fixed part which we want to keep from the original test configuration - s_{fixed} , the fault type which we current focus - F_i . And the values sets that each option can take from respectively. The output of this algorithm is a test configuration t_{new} which either trigger the expected F_i or passed.

In fact, this algorithm is a big loop(line 1 - 14) which be parted into two parts:

The first part(line 2 - line 7): generate a new test configuration which is different from the original one. This test configuration will keep the fixed part (line 7), and just mutant these factors are not in the fixed part(line 2). The mutant for each factor is just choose one legal value that is different from the original one(line 3 - 6). The choosing process is just by random and the generated test configuration must be different each iteration(can be implemented by hashing method).

Second part is to validate whether this newly generated test configuration matches our expect(line 8 - lone 13). First we will execute the SUT under the newly generated test configuration(line 8), and then check the executed result, either passed or trigger the expected fault - F_i will match our expect.(line 9) If so we will directly return this test configuration(line 10). Otherwise, we will repeat the process(generate newly test configuration and check)(line 11 -12).

It is noted that the loop have another end exit besides we have find a expected test configuration(line 10), which is when function *MeetEndCriteria()* return a true value(line 1). We didn't explicitly show what the function *MeetEndCriteria()* is like, because this is depending the computing resource you own and the how accurate you want to the identifying result to be. In detail, if you want to your identify process be more accurate and you have enough computing resource, you can try much times to get the expected test configuration, otherwise, you may just try a relatively small times to get the expected test configuration.

In this paper, we just set 3 as the biggest repeat times for function. When it ended with *MeetEndCriteria()* is true,

Table 9: Identifying MFS using OFOT with our approach

original test configuration	fault info
0 0 0 0	Err 1
gen test configurations	result
1 0 0 0	Err 1
0 1 0 0	Err 2
<u>0 2 0 0</u>	Err 1
0 0 1 0	Pass
0 0 0 1	Pass
original identified:	identified with re- placement
(- 0 0 0)	(- - 0 0)

we will return null(line 15), which means we cannot find a expected test configuration.

5.2 Examples when apply this approach into some MFS identifying algorithms

Next we will take three MFS identifying algorithms as the subject to see how our approach works on them.

5.2.1 OFOT examples

Assume we have test a system with four parameters, each has three options. And we take the test configuration (0 0 0 0) we find the system encounter a failure called "Err 1". Next we will take the MFS identifying algorithms – OFOT with the help of our approach to identify the MFS for the "Err 1". The process is listed in table 9. In this table, The test configuration which are labeled with a deleted line represent the original test configuration generated by OFOT, and it will be replaced by the regenerated test configuration which are labeled with a wave line under it.

From this table, we can find the algorithm mutant one factor to take the different value from the original test configuration on time. Originally if the test configuration encounter the different condition with the Err 1, OFOT will make a judgement that the MFS was broken, in another word, if we change one factor and it does not trigger the same fault, we will label them as one failure-inducing factor, after we changed all the elements, we will get the failure-inducing schemas. For this case, as when we change the second factor, third factor and the fourth factor, it doesn't trigger the Err 1 (for second factor, it trigger Err 2 and for the third and fourth, it passed). So if we do not regenerate the test configuration (0 2 0 0), we will get the MFS – (- 0 0 0) for the err 1(which are also labeled with a delete line).

However, if we replace the test configuration (0 1 0 0) with (0 2 0 0) which triggered err 1 (in this case, the fixed part of the test configuration is (0, - - -)), we will find that only when we change the third and fourth factor will we broke the MFS for err 1, so with our approach, we will find the MFS for err 1 should be (- - 0 0) (labeled with a wave line under it).

5.2.2 CTA examples

CTA uses the covering array as its inputs and then use classification tree algorithm to characterize the MFS. For

Table 10: Identifying MFS using CTA with our approach

covering arrays	fault info
0 0 0 0	Err 1
0 1 1 1	Err 2
<u>1 1 1 1</u>	Err 1
<u>0 1 1 0</u>	Pass
<u>0 0 0 1</u>	Pass
0 2 2 2	Pass
1 0 1 2	Pass
1 1 2 0	Err 1
1 2 0 1	Err 1
2 0 2 1	Pass
2 1 0 2	Pass
2 2 1 0	Pass
original identified:	identified with re- placement
(0 0 - -) : err1	(0 0 0 -):err1
(1 1 - -) : err1	(1 1 - -):err1
(1 2 - -) : err1	(1 2 - -):err1
(0 1 - -) : err2	

this algorithm, we still assume the SUT has 4 parameters and each one has 3 values. Then we will initial a 2-way covering array as the input for CTA algorithm which is listed in table 10. The delete line and wave line have the same meaning as the OFOT example.

Let's look at the original test configuration (0 1 1 1), it triggered the err 2 which is not as our expected, so we will replace it with other test configurations. A prerequisite for this replacement is that we should not decrease the covering rate. As the original test configuration contain the 2-degree schema(0 1 - -),(0 - 1 -),(0 - - 1),(-,1,1,-),(-,1,-,1),(-,-,1,1). We make them as fixed part that the regenerate test configuration must keep, as we cannot use one test configuration to cover all the six fixed part, so instead, we use three additional test configurations (1 1 1 1), (0, 1, 1, 0) and (0,0,0,1) to cover them, note that this three test configurations either trigger err 1 or pass. If they don't match this condition, we will try other test configurations to replace the original test configuration.

5.2.3 FIC_BS examples

FIC adaptively generate test configurations according to the executing result of last test configuration. So we will observe each time it generate a test configuration to see whether should we replace the newly generated test configuration. Take the following example, that a SUT contain 8 parameters and each parameter has three values, we set the number of parameters to be 8 because it can get a better description of this algorithm.

As the last test configuration will have a significance impact for FIC, so if we replace one test configuration of the FIC_BS, the generated test configurations will have a big difference from traditional FIC_BS. To clearly describe the influence of our approach on FIC_BS, we will first give a

completed example of the traditional identifying process of FIC_BS, and then give the process with our approach. The detailed is listed in table 11.

For the first part, we can easily learn that FIC_BS use a binary search strategy to mutant the parameters in a test configuration. And which part will be mutant is based on the last test configuration. We will not going to describe the FIC_BS algorithms in detail. Instead, we just focus on the test configuration (1 1 1 0 0 0 0 0) trigger err 2. As traditional FIC_BS will regard it as a passing test configuration, take the previous test configurations together, it will judge that (- - 0 - - - -) must be a failure-inducing factor. And finally it will take (- - 0 0 - 0 - -) as the MFS for err 1.

However, when apply our approach, we will replace the test configuration (1 1 1 0 0 0 0 0) to be (2 2 2 0 0 0 0 0), ((- - 0 0 0 0 0) is the fixed part). And then we will find it still trigger the err 1, which means (- - - 0 - - -) should be failure-inducing factor rather than (- - 0 - - - -) in the traditional approach. And this step will have a influence on the test configurations generated following, and finally we will identify the schema (- - - 0 - 0 - -) as the MFS for err 1.

6. EMPIRICAL STUDIES

We conducted several case studies to address the following questions:

Q1: What the condition of faults and their masking effect in real software.

Q2: What the behaviour of traditional approaches works on software when they have multiple faults?

Q3: Does our approach get a better solution than traditional approaches when facing multiple faults?

Specifically, section 6.1 survey several open-source software to gain a insight of the condition of multiple faults and their masking effects. Section 6.2 applied three MFS-identifying programs on the faults-controllable software from the SIR benchmark. Section 6.3 apply our approach on the benchmark and a comparison with traditional approaches will be discussed. Section 6.4 discuss the threats to validity of our empirical studies.

6.1 study 1: multiple faults and masking effects in practice

In the first study, we survey several software to gain a insight of the condition of multiple faults and their effects. The software under study are four GNU software: Grep, Sed, Make, Gzip and a database management software- H-SQLDB. Each of them contain different versions. The reason why we choose them as the subject is that these software are highly configurable, the options and their combination can influence their behaviour, additionally, they all have the developer community so that we can easily get the real bugs reported in the bug tracker forum. Table 12 lists the program, the number of versions we surveyed, number of lines of uncommented code, number of procedures(for c software) or number of classes (for java software) and a brief description.

6.1.1 study setup

For each software listed in the table 12, we checked its bug tracker system and focus on the bugs related to the versions we cared. We collected the bugs to each version of the software and exam whether they can masked each other in some test cases. In detail, we will exam each pair of the

Table 11: Identifying MFS using FIC_BS

original test configuration	fault info
0 0 0 0 0 0 0 0	Err 1
Use traditional FIC_BS	
gen test configurations	result
1 1 1 1 0 0 0 0	PASS
1 1 0 0 0 0 0 0	Err 1
1 1 1 0 0 0 0 0	Err 2
1 1 0 1 1 1 1 1	PASS
1 1 0 1 1 0 0 0	PASS
1 1 0 1 0 0 0 0	PASS
1 1 0 0 0 0 0 0	Err 1
1 1 0 0 1 1 1 1	PASS
1 1 0 0 1 1 0 0	PASS
1 1 0 0 1 0 0 0	Err 1
1 1 0 0 1 0 1 1	Err 1
Mfs identified	
(- - 0 0 - 0 - -) for Err 1	
Use FIC_BS with our approach	
gen test configurations	result
1 1 1 1 0 0 0 0	PASS
1 1 0 0 0 0 0 0	Err 1
1 1 1 0 0 0 0 0	Err 2
<u>2 2 2 0 0 0 0 0</u>	Err 1
1 1 1 0 1 1 1 1	PASS
1 1 1 0 1 1 0 0	PASS
1 1 1 0 1 0 0 0	Err 1
1 1 1 0 1 0 1 1	Err 1
Mfs identified	
(- - - 0 - 0 - -) for Err 1	

Table 12: Software under survey

software	versions	LOC	procedur/classes	description
Grep	5	9493-10102	146	a program to search for strings inside a file.
Sed	7	5503-14477	255	A stream editor that parses and transforms text
Make	4	13359-35583	268	a utility that automatically builds executable programs and libraries.
Gzip	5	4604-5754	104	a popular data compression program
HSQldb	2	-	-	A database management software written with pure java

Table 13: number of faults and their masking effect condition

software	versions	Num of faults	Num of masking
Grep	-	49	-
Sed	-	-	-
Make	-	131	-
Gzip	-	-	-
HSQldb	-	-	-

bugs we collected for a particular version, and then check whether they can fulfil the following condition:

1) Can they exist in one test case? 2) Does only one fault can be observed in the test case output?

The first condition will ensure that the options or inputs that trigger this two bugs is compatible. For example, if bug₁ need the input $a = 0$, and bug₂ need the input $a = 1$, then this two bug will not exist in one test case. The second condition ensure that the masking effect can appear, as if only one fault information can be observed, then another fault will be ignored in the test case.

The number of masking effect among the bugs for a specific version will be given in the next section.

6.1.2 result and discuss

Table 13 lists the results of our survey. Column "Num of faults" indicate the number of faults we collected from the bug tracker system, column "num of masking" indicate the number of fault pairs which fulfil the aforementioned two conditions.

The detailed fault information like the fault ID, masking fault ID and its posted web site have been attached in the Appendix A, table 16. All of these bug can be rebuilt by following the steps reported in the bug tracker. For example, Bug-#29537 indicate that Grep incorrectly match unicode patterns with $\langle \rangle$ and Bug-#33080 claim that combination of $-count$ ($-c$) and $-only-matching$ ($-o$) will not work as expected. A simple test case liked the following command:

```
grep -only-matching -count '^<[^\>]*>' test.txt
```

can make these two bugs existed in one test case and only one fault will be observed. Other faults listed in table 16 as well as their masking effect can also be easily reappeared in well-designed test cases.

From table 13, we can find for a particular version of software, multiple faults can be discovered, and these faults can easily existed in one test case which can let the masking effect triggered. So the answer to Q1 is We can see the multiple faults is a common sense and the masking effect do exist in real software.

Table 14: subjects model

software	Model
Grep	$3*4*5$
Sed	$3*4*5$
Make	$3*4*5$
Gzip	$3*4*5$
HSQldb	$3*4*5$

6.2 performance of traditional algorithms

Previous section tell us multiple faults do exist in software and they may trigger masking effect. So in this section we want to learn how they impact on the traditional approached. To get this target, we need to do two group experiments, first, what the behaviour of these approaches worked on software with single fault, second what the behaviour of these approaches worked on multiple faults. So the ideal subject for this case study must be the software which can easily controlled with faults injected. To get this target, we use the SIR benchmark as our subject and one real software as the subject. We model them as myra cohen do[] to get a combinatorial testing model. The detailed model is listed as table 14. And the HSQldb can be modeled as table 17

6.2.1 study setup

For each version of software, we do the following: First use the TSL describe the model of software, generate test cases and run them. Remodel them as Cohen do. The select the failing test cases, and use algorithms to identify the result. And then use multiple faults, do the same as previous said, then compared the results as the followed

6.2.2 result and discuss

6.3 performance of our approach

To evaluate our approach, we take two open-source software as our subjects: GNU software: Grep, Sed, Make, and so on. HSQldb and Log4j. HSQldb is a pure-java data-base management software, and Log4j is a logger tool commonly used in java application. Both of them have a large support developers' forum. Furthermore as our approach is a framework that can be seen as a for the MFS identifying algorithms, so that we afford them five algorithms: fic ofot trt cta sp.

6.4 threats to validity

6.5 experiment set up

We will first model the two objects so that we can use the MFS identifying algorithms. To do this, we first find real faults in the developer's forum, and then read the specification of the software to get the options and its values that can influence its behaviour. The detail of the modeling is listed in table 17 and table .

Second we will write the test script that can autonomically execute the software according to a given test configuration, furthermore, the test script will collect the executing result. In this paper, we just care two kinds of result: the test script run without any exception and run with triggering exception. For the second result we will distinguish the failure with different exception information.

Third we will take three group experiments, each for the question proposed in the first of this section.

For the first experiment, we will compare the result got by traditional MFS identifying algorithm with the MFS identifying algorithms using our framework. For a particular algorithm, say, OFOT, we will record the number of MFS, the degree of MFSs and the number of test configurations they needed.

The second experiment, we will collect the result got by the algorithms with our framework. Then we will generate a test configuration contain two MFS of the result, and execute to find if one MFS is masked by another. In specific, we will exam every possible pair MFSs, the only constraints is that this two pair MFS must get different fault information.

For the third experiment, we will compare the similarity between the identified result and the result report in the developer's forum. And find whose (traditional or with our framework) result can give more help to the developer of the software.

6.6 evaluation metrics

To evaluate the efficiency of our framework, we will compare the results of two approaches. We will count how many MFSs is reduced by our approach and how the degree is reduced by our approach compared to traditional one. This two metric will indicate how much our approach will improve the usability of results. Additional, we will count how many more test configurations we will generated compare to traditional one, this metric indicate the cost we need over the traditional one.

Then we will exam the result of the algorithm with our framework, to see if any masking is behind them. In detail, we will record how many pair MFS can coexist in a test configuration. And count the privileges of each MFS.

For the third, the similarity is listed as follows: the followed notation: Assume we get two schema S_A, S_B , the notation $S_A \cap S_B$ indicate the same elements between S_A and S_B . For example $(-1 \ 2 \ -3) \cap (-2 \ 2 \ -3) = \{(-2 \ -2 \ -), (- \ - \ - \ 3)\}$.

$$Similarity(S_A, S_B) = \frac{|S_A \cap S_B|}{\max(Degree(S_A), Degree(S_B))}$$

In this formula, the numerator indicate the number of same elements in S_A and S_B . It is obviously, the bigger the value is the more similar two schemas are. The denominator give us this information: if the the number of same elements is fixed, the bigger the degree of the schema, the more noise we will encounter to find the fault source. In a word, the bigger the formula can get the better the similarity is between

Table 15: Masking effect state

all pair	coexisted & noexisted	mask over 6 mfs	over 5 mfs
10	5 & 5	3	2

the two schema.

6.7 result and analysis

For the second experiment.

7. RELATED WORKS

combinatorial testing has many factors, Nie give a survey, at some are focus,.

Identify MFS are

then consider the masking effect is , to my best knowledge, only one paper, unfortunately, it just consider the .

Yilmaz propose a work that is feedback driven combinatorial testing, different from our work, it first using CTA classify the possible MFS and then eliminate them and generate new test cases to detect possible masked interaction in the next iteration. The difference is that the main focus of that work is to generate test cases that didn't omit some schemas that may be masked by other schemas. And our work is main focus on identifying the MFS and avoiding the masking effect.

8. CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the L^AT_EX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

9. ACKNOWLEDGMENTS

This section is optional; it is a location for you to acknowledge grants, funding, editing assistance and what have you. In the present case, for example, the authors would like to thank Gerald Murray of ACM for his help in codifying this *Author's Guide* and the .cls and .tex files that it describes.

10. ADDITIONAL AUTHORS

Additional authors: John Smith (The Thørväld Group, email: jsmith@affiliation.org) and Julius P. Kumquat (The Kumquat Consortium, email: jpkumquat@consortium.net).

11. REFERENCES

APPENDIX

A. MASKING FAULTS STATISTIC

B. HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are different in the appendices. In the **appendix** environment, the command **section** is used to

Table 16: real faults detailed

software	faults ID	masking faults ID	Web site
Grep	#29537	#33080	http://savannah.gnu.org/bugs/?group=grep
-	#29537	#33080	-
Sed	-	-	-
Make	-	-	-
Gzip	-	-	-
HSQLDB	-	-	-

B.1 References

Generated by bibtex from your .bib file. Run latex, then bibtex, then latex twice (to resolve references) to create the .bbl file. Insert that .bbl file into the .tex source file and comment out the command `\thebibliography`.

C. MORE HELP FOR THE HARDY

The sig-alternate.cls file itself is chock-full of succinct and helpful comments. If you consider yourself a moderately experienced to expert user of L^AT_EX, you may find reading it useful but please remember not to change it.

Table 17: input model of HSQLDB

SQL properties(TRUE/FALSE)	
sql.enforce_strict_size, sql.enforce_names,sql.enforce_refs, sql.enforce_size, sql.enforce_types, sql.enforce_tdc_delete, sql.enforce_tdc_update	
table properties	values
hsqldb.default_table_type	CACHED, MEMORY
hsqldb.tx	LOCKS, MVLOCKS, MVCC
hsqldb.tx_level	read_committed, SERIALIZ- ABLE
hsqldb.tx_level	read_committed, SERIALIZ- ABLE
Server properties	values
Server Type	SERVER, WEBSERVER, IN- PROCESS
existed form	MEM, FILE
Result Set properties	values
resultSetTypes	TYPE_FORWARD_ONLY,TYPE_SCROLL_INSENSITIVE, TYPE_SCROLL_SENSITIVE
resultSetConcurrencys	CONCUR_READ_ONLY,CONCUR_UPDATABLE
resultSetHoldabilitys	HOLD_CURSORS_OVER_COMMIT, CLOSE_CURSORS_AT_COMMIT
option in test script	values
StatementType	STATEMENT, PREPARED- STATEMENT