Test Case Generation and Optimization for User Session-based Web Application Testing

Qian Zhongsheng

School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China
Email: changesme@163.com

Abstract—**Web application testing is the process of revealing errors that is used to give confidence that the implementation of a Web application meets its original specification. An approach to reducing and optimizing the test cases generated from user request traces is presented. A large volume of meaningful user sessions are obtained after purging their irrelevant information by analyzing user logs on the Web server. Most of the redundant user sessions are also removed after reducing them. For test reuse and test concurrency, it divides the user sessions obtained into different groups, each of which is called a test suite, and then prioritizes the test suites and the test cases of each test suite. So, the initial test suites and test cases, and their initial executing sequences are achieved. However, the test scheme generated by the elementary prioritization is not much approximate to the best one. Therefore, genetic algorithm is further employed to optimize the results of grouping and prioritization. Meanwhile, an approach to generating new test cases is presented using crossover.**

Index Terms—**user session, genetic algorithm, test case, test suite, reduction, prioritization, grouping, common prefix**

I. INTRODUCTION

Web applications have been closely linked to health, economies, finance, sales, retail, marketing, etc. Their failure would cause great inconvenience or even may result in great disaster. Therefore, a Web application must be reliable enough and meets its original user specifications. Web testing is one of the methods to ensure the quality of Web applications.

Most original Web testing methods are repetitive and inefficient, as results in low reliability and high risk in developing Web applications. Many aspects regarding Web testing have not been sufficiently investigated yet, and many open questions still need to be addressed. This work investigates a key problem in Web application testing: test case generation and optimization[†]. It delves into an approach to testing and optimizing Web applications based on user sessions using genetic algorithm. When converting a user session to a corresponding test case, we preserve the user input data.

II. USER SESSION-BASED TESTING

User session-based testing is an automated approach to enhancing an initial test suite with actual user data, enabling additional testing during maintenance as well as adding test data that represents usage as operational profiles evolve [1]. When a new test case is added to the test suite to test a new or changed requirement, or perverse the adequacy of the test suite, the size of the test suite increases, and the cost of running the test suite in a modified Web application increases too.

In user session-based testing, each user session begins when a user from a new IP address makes a request from the server and ends when the user leaves the Web site or the session times out, i.e., a user session is a collection of user requests in the form of URL and name-value pairs. To convert a user session into a test case[†], each logged request is changed into an HTTP request that can be sent to a Web server. A test case consists of a set of HTTP requests that are associated with each user session. Different conversion strategies from user sessions to test cases were proposed [2-3].

It often takes the testers too much to run all the tests in the test suite. The researchers have presented two methods of keeping the same coverage ratio as the original test suite, to investigate the effectiveness[†] and performance of a test suite. These two methods are the reduction of test suites [4-6] and the prioritization of test suites [7-8].

Reducing test suites is critical to the user session-based testing. Generally, the reduction method considers each test case one by one, discards those that do not change or make an influence on the test requirements, and guarantees that the resulting test suites preserve the same properties as the original test suites, such as the same test coverage ratio or fault detection capabilities.

Zhang, et al. [9] proposed a test requirement-based reduction model that can describe the interrelations among test requirements precisely. Based on the model, they presented a test requirement-based reduction method to generate the reduced test requirement set, which is the basis of test suite generation, reduction and optimization. Sprenkle, et al. [1] proposed another requirement-based reduction technique, which processes in a cut-and-try method.

[†] Optimization means the process of grouping and prioritizing test suites and test cases generated using genetic algorithm.

[†] Generally, we do not distinguish user session from test case unless otherwise specified, for each user session processed is converted into a test case, i.e., there is a one-to-one relationship between them directly.

[†] A test suite with higher effectiveness can satisfy test requirements given as soon as possible.

Compared to the reduction of test suites, the prioritization, which identifies the sequence of test suites according to some criterion, adds the strongest test case into new test suite. That's to say, a test case, which plays a most important role on the coverage ratio of test requirements is added. The test cases of higher priorities are executed earlier than those of lower priorities using the prioritization technique to satisfy some requirements as soon as possible. For example, testers may hope to increase code coverage ratio for a Web application under test more rapidly. They may also hope to enhance their confidence in the reliability of a Web application earlier. It is evident that the prioritization technique itself does not discard any test case.

If the testing must be terminated earlier, the prioritized test suites provide higher efficiency [†] in detecting faults than those test suites, which are not prioritized. Intuitively, if the probability of a test suite is higher than others in detecting faults in a given time, then it has better fitness.

In addition, test selection technique is also relevant to the effectiveness and performance of test suites [10]. A subset of test suites is used to process all the changes. Generally, these test suites, however, can not provide the same coverage ratio as the original test suites. This method can be regarded as an incremental testing.

III. TEST CASE GENERATION AND OPTIMIZATION

User session-based testing reflects the actual usage situation, as can not be predicted in the early development stages. The user session data acquired from actual users are a useful supplement for the testers to generate test suites in-house [2]. We conduct user session-based Web application testing using genetic algorithm.

A. Collecting and Reducing User Sessions

In the logs on each Web server, each access record corresponds to a request by a user each time. The contents of the record include request source (user IP address), request time, request mode (such as GET, POST), the URL of requested information, data transport protocol (such as HTTP), status code, the number of bytes transferred and the type of client, etc. It needs to scan the logs only once to resolve the original active historic records from current access logs. However, it is difficult to organize these original records directly, which must be preprocessed. We first purge irrelevant data including the records whose status codes are erroneous (the code 200 for success, 400 for error), embedded resources such as script files and multimedia files whose extension names are .gif, .jpeg or .css, etc., to obtain the set of user sessions for primitive analysis. Then, we create user sessions through scanning the logs on Web servers. Once a new IP address occurs, a new user session is created. The sequential requests sent from this IP address are appended to the new session under the

condition that the time interval of two continuous requests is not greater than max-session-idle-time pre-determined, or else another new user session begins. The set of all the user sessions is finally achieved.

There are often large volumes of user sessions collected and a user session is converted to a test case transmitted to Web server. Therefore, we can eliminate *redundant* user sessions using reduction techniques, and then preserve *necessary* user sessions. For the convenience of discussion, some important concepts are given first.

Definition 1 (URL trace)   A *URL trace* is the URL sequence requested by a user session.

Let α be a URL trace. Its length is the number of URLs requested in the trace, denoted by |α|.

Definition 2 (prefix)   A trace α is the *prefix* of another trace β, if and only if α is the subsequence of β and they have the same initial symbol.

Definition 3 (common prefix)   If a trace is the prefix of several traces, then this trace becomes their *common prefix*.

Definition 4 (greatest common prefix)   The longest common prefix in all the common prefixes of two traces is their *greatest common prefix*.

The longer the greatest common prefix of two traces is, the more similar the two traces are, i.e., their similarity is higher. Let we have four traces that are γ_1=abcdefg, γ_2=abcdeh, γ_3=abcd and γ_4=cde. Then, γ_3 is the common prefix of γ_1 and γ_2, but not the greatest one. The greatest common prefix of γ_1 and γ_2 is abcde. Although γ_4 is the subsequence of γ_1 and γ_2, it is not their prefix, for the initial symbol of γ_4 is not the same as that of γ_1 and γ_2. We define a function isPrefix(α, β), which decides whether or not α is the prefix of β, i.e., whether or not the URL trace requested by a user session is *redundant* corresponding to that requested by another user session. If there is a prefix, then the function returns a BOOLEAN value TRUE, or else returns FALSE. The URL trace-based user session algorithm ReduceUSession is shown in Fig. 1. Note that, an HTTP request here is regarded as a symbol of a trace.

The ReduceUSession algorithm decides whether or not a URL trace α requested by a user session is the prefix of β requested by another user session. If it is true, then the user session corresponding to α is removed. The number of user sessions obtained using this algorithm will be reduced greatly.

```
Algorithm: ReduceUSession
 input:
       The set of user sessions Λ={s₁, …, sₖ}, where k is the number of
       user sessions;
       The URL trace U₁, …, Uₖ, which are requested by s₁, …, sₖ
       respectively;
 output:
       The reduced set of user sessions denoted by Γ;
 begin
       Γ=Φ;
       while (another user session that is not marked in Λ exists)
            tag1=FALSE;
            tag2=FALSE;
            Select a user session sᵢ that is not marked in Λ, and then
            mark it with "USED";
            for (the URL trace Uⱼ requested by each user session sⱼ in Γ)
```

---

† Higher efficiency means less execution time or less test times to satisfy the same test requirements.

```
        if isPrefix(U_j, U_i)      //the URLs requested by s_i is
                                   //more, so s_j is redundant
            Γ=Γ-{s_j};
            tag1=TRUE;
        endif;
        if isPrefix(U_i, U_j)      //here, Γ keeps unchanged, and
                                   //s_i is redundant
            tag2=TRUE;
            break;                 //exit for cycle
        endif;
    endfor;
    if tag1 || (!tag1 && !tag2)    //s_i is necessary
        Γ=Γ ∪ {s_i};
    endif;
  endwhile;
  Output the reduced set of user sessions Γ;
end.
```

Figure 1. The user session reduction algorithm ReduceUSession

The ReduceUSession algorithm is different from other user session reduction algorithms. Most other reduction algorithms analyze each test case in the test suites one by one and remove those test cases that cannot change or affect test requirements, i.e., they distinguish *redundant* and *necessary* test cases, as is too difficult to manage in practice, for it is very intractable to discriminate that the test requirements satisfied by some test cases (i.e., redundant ones) are also satisfied by other test cases before the test execution using the existential algorithms. While our ReduceUSession algorithm decides whether or not a URL trace requested by a user session is the prefix of another URL trace requested by another user session. Based on this, we can identify the *redundant* test cases, as can be easily done in practice. Besides, it covers all the URLs requested by the original set of user sessions and keeps the sequence of URL requests, i.e., it guarantees that the original test requirements are satisfied.

B. Grouping and Prioritizing User Sessions

The user sessions reduced by the ReduceUSession algorithm are divided into subgroups, each of which is regarded as a test suite. The goal of grouping is to reuse test cases and the testing can be executed at different platforms in parallel (or concurrently), to lessen test time and improve test efficiency. Moreover, the interacting test can be conducted between a pair of user sessions in each group. We try our best to keep the property for the user sessions in the same group that the URL traces requested bear greatest common prefix of a certain length. Several discontinuous integral threshold values, denoted by $\zeta_1, \zeta_2, \dots \zeta_k$ ($\zeta_i \geq 1$, $1 \leq i \leq k$), are defined. The user sessions, the lengths of whose greatest common prefix of URL traces requested fall in between a certain threshold values, are grouped together. Fox example, let we have three threshold values $\zeta_1=2$, $\zeta_2=4$ and $\zeta_3=7$, then the user sessions are divided into four groups that are S_1, S_2, S_3 and S_4, the lengths of whose greatest common prefix (denoted by α) are $|\alpha| \leq 2$, $2 < |\alpha| \leq 4$, $4 < |\alpha| \leq 7$ and $|\alpha| > 7$ respectively. The test cases in these four groups can be executed at different platforms in parallel (or concurrently), and the greatest common prefix is also reused in each group. Notice that this is not the unique grouping way. For example, in the four traces $\gamma_1=abcdefg$, $\gamma_2=abcdeh$, $\gamma_3=abcd$ and $\gamma_4=cde$, we can divide them into two groups $\{\gamma_1, \gamma_2\}$ and $\{\gamma_3, \gamma_4\}$, the lengths of whose greatest common prefix are $4 < |\alpha| \leq 7$ and $|\alpha| \leq 2$ respectively, or another two groups $\{\gamma_1, \gamma_2, \gamma_3\}$ and $\{\gamma_4\}$, the lengths of whose greatest common prefix are $2 < |\alpha| \leq 4$ and $|\alpha| \leq 2$ respectively. Of course, it is unnecessary to require that the greatest common prefix of URL traces requested by any two user sessions in each group fall in between a certain threshold values; it is recommended that most of them satisfy this property (a percentage can be pre-designed or generated randomly for the measurement). A compromise way needs to be found to classify these URL traces (or test suites), i.e., a tradeoff should be found between grouping and concurrent testing and test reuse. Suppose that N user sessions are divided into K groups. In general, there is almost the same number of user sessions in one group as that in another, i.e., the number equals approximately to $\left\lceil \frac{N}{K} \right\rceil$. Additionally, this way of grouping is preparatory, called *elementary grouping*.

The common prefix indicates the users' common events, or the same or similar operations. It also shows that the users bear the same or similar interests. The longer the common prefix is, the more evident it is, as is the case of most users. In addition, there is a special group of user sessions, the length of whose greatest common prefix of URL traces requested is shortest. This group of user sessions often indicates different URL requests, which represent distinct requirements for a Web application. In these sessions, many aberrant events often occur with unwonted input data. They belong to boundary cases, which are very easy to go wrong for the Web application.

Herein, we prioritize test suites. The test suite with shortest length of common prefix ranks first, then all the other test suites are arranged according to their lengths of common prefix in descending order. So, the test suite in final position is that whose length of common prefix is last but one. In the test suites S_1, S_2, S_3 and S_4, the lengths of whose common prefixes are $|\alpha| \leq 2$, $2 < |\alpha| \leq 4$, $4 < |\alpha| \leq 7$ and $|\alpha| > 7$ respectively, if we prioritize them, then the test executing sequence for those test suites are S_1, S_4, S_3, S_2. That is to say, the test cases in S_1 are executed first, then the test cases in S_4 and S_3 are executed respectively and finally, the test cases in S_2 are executed. In each test suite, the test cases are prioritized according to the coverage ratios of URLs requested, i.e., the test case with longer URL trace requested is executed earlier. If the lengths of URL traces of several test cases in the same test suite are equal, then they are randomly executed.

Our approach of prioritization is different from those presented before. The existing methods of prioritization add the strongest test case, which is of the maximal use for the coverage ratio of test requirements, to the new test suite. Those methods aim to execute earlier the test cases of high priority than those of lower priority, to satisfy some test requirements as soon as possible. These methods, however, are often difficult to find the (nearly)

strongest test case each time before test run, while our approach divides test cases into several groups according to the idea of common prefix before prioritizing them. It is more convenient for the testers to selectively execute the test cases of some group by grouping all the test cases first, to detect some types of faults, for the same or similar types of errors are often detected by those test cases in the same groups. In addition, test cases can be reused by grouping and the testing can be executed at different platforms in parallel (or concurrently), to lessen test time and improve test efficiency. Moreover, we have considered a special type of user sessions, the length of whose greatest common prefix of URL traces is shortest. This group of user sessions often contains specific requests, which are the primary source for errors.

C. Testing Web Applications Using Genetic Algorithm

After the process of grouping and prioritization above, we obtain several initial test suites and test cases with the initial executing sequences. However, the test scheme generated by the elementary prioritization is not fast in finding faults and can not satisfy the requirements earlier. Therefore, genetic algorithm is employed further to optimize the grouping and prioritization.

In 1975, an American professor Holland first proposed the idea of genetic algorithm systematically [11]. It has attracted a large number of researchers and extended into those aspects of optimization, search and machine learning with a solid theoretic foundation. The genetic algorithm, a global optimization search algorithm of high efficiency, focuses on all the individuals in one population, and uses random techniques to search efficiently for a coded parameter space. Selection, crossover and mutation are the basic operators in genetic algorithm. And parameter coding, the setting of initial population, the design of fitness function, the selection of genetic operations and control parameters consist of the critical part of genetic algorithm.

In the following, we test Web applications using genetic algorithm to further optimize the initial test suites and test cases, and their initial executing sequences, in order to achieve better test suites and test cases that satisfy test requirements. The process of yielding a population of next generation using three basic operators that are selection, crossover and mutation once is called *an iteration*. To obtain a good result, much iteration is repeated. The following introduces the process of selection, crossover and mutation.

1) Selection

A pair of individuals is selected from a parent population with the probability of p_s. The probability that an individual is selected is in direct proportion to its fitness value, as is often implemented using the strategy of roulette wheel [11]. In selecting, the individual of high fitness value is duplicated into the population of next generation directly. The higher the fitness value of an individual is, the higher the probability of yielding its offspring is, as shows that it is more appropriate to the expected result. Let we get K test suites (constituting the initial population), which are S_1, S_2, ..., S_K of the descending order according to the prioritization

technique. In practice, we combine the error coverage ratios of test suites and the cost of test run to design fitness function. The fitness value is listed as f_1, f_2, ..., f_K from high to low, where f_i is the fitness of S_i ($1 \le i \le K$). The probability that an individual is selected equals to the resulting value that its fitness value divides the sum of fitness values of all the individuals, i.e., the selected probability of S_i, denoted by p_s^{Si}, is $f_i \Big/ \sum_{j=1}^{K} f_j$. Obviously, the sum of the selected probabilities of all the K test suites equals to 1.

According to the discussion above, the fitness f_1 and f_2 corresponds to two special test suites S_1 and S_2, the lengths of whose greatest common prefixes are shortest and longest respectively. S_1 and S_2 are selected to be the individuals of next generation directly (in practical use, we can also select more than two individuals of high fitness to become the next individuals); in case they do not be selected. Now, we randomly yield two numbers that are g_1, $g_2 \in [0, 1]$, and randomly select two test suites that are S_i and S_j whose probabilities are not less than g_1 and g_2 respectively. The two new test suites S_i' and S_j' (the individuals of next generation) are generated through crossovering and mutating their parents S_i and S_j. Repeat the process of selection, until adequate test suites of next generation are yielded. One point should be emphasized that some test suites of higher probabilities may not be selected as parents, while others of lower probabilities are selected. This case is reasonable and accords with the theory of biological evolution in nature, because any thing has its necessity and occasionality at the same time.

2) Crossover

The genes chain of two parent individuals selected are crossovered with the probability of p_c using the TSCrossover algorithm, where p_c is a system control parameter. The new individuals after crossovering are used to replace their parents. The TSCrossover algorithm is shown in Fig. 2.

```
Algorithm: TSCrossover
input:
      Parent test suites Si and Sj;
      The crossover probability of pc;
output:
      Si' and Sj' of next generation;
begin
      Si', Sj'←Si, Sj;
      Randomly generate a number g∈[0, 1];
      if pc is not less than g
            Randomly generate an integral number g',
            g' ∈(0, min(|Si'|, |Sj'|)];
            Interchange the test cases in Si' and Sj' from the position of
            g' one by one, until no more test case needs to be
            interchanged in any of them;
      endif;
      Output Si' and Sj';
end.
```

Figure 2. The crossover algorithm TSCrossover for test suites

Let we have two test suites $S_i = \langle c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9 \rangle$, which contains 9 test cases and $S_j = \langle c_{10}, c_{11}, c_{12}, c_{13}, c_{14}, c_{15}, c_{16} \rangle$, which contains 7 test cases. If the position of crossovering is 3, then two new individuals of

next generation after crossovering S_i and S_j are: $S_i'=<c_1$, c_2, **c_{12}, c_{13}, c_{14}, c_{15}, c_{16}**, c_8, $c_9>$ and $S_j'=<c_{10}$, c_{11}, **c_3, c_4, c_5, c_6, $c_7>$**, where the test cases indicated in bold face are those interchanged one by one from corresponding test cases in parent test suites.

3) Mutation

Each bits of the genes chain of new individuals are mutated with the probability of p_m using the TSMutation algorithm, where p_m is a system control parameter. The probability of mutation is often small, or else it will raise questions over the system stability. The populations can be prevented from stagnating through mutation. If there is no mutation, then the test data of new populations will be confined to the initial values. The mutation process is conducted respectively using the TSMutation algorithm for the test suites S_i' and S_j' crossovered using the TSCrossover algorithm. The new individuals after mutating are employed to replace those before mutating. The TSMutation algorithm is shown in Fig. 3.

```
Algorithm: TSMutation
input:
    The test suites S_i' and S_j' after using TSCrossover algorithm;
    The mutation probability of p_m;
Output:
    The test suites S_i' and S_j' after mutating;
begin
    for each S in {S_i', S_j'}
        for each c in S
            Randomly generate a number g∈[0, 1];
            if p_m is not less than g
                Interchange the positions of c and the test case
                before c;
            endif;
        endfor
        Output S;
    endfor
end.
```

Figure 3. The mutation algorithm TSMutation for test suites

Let $p_m=0.015$ and $S_i'=<c_1$, c_2, c_{12}, c_{13}, c_{14}, c_{15}, c_{16}, c_8, $c_9>$, which is one of the test suites after using the TSCrossover algorithm. We randomly generate some data: $g(c_1)=$ 0.246, $g(c_2)=$ 0.580, $g(c_{12})=0.025$, $g(c_{13})=0.012$, $g(c_{14})=0.632$, $g(c_{15})=$ 0.073, $g(c_{16})=0.431$, $g(c_8)=0.193$ and $g(c_9)=0.059$. For p_m is not less than $g(c_{13})$, c_{13} is mutated, i.e., the positions of c_{13} and c_{12} are interchanged. The test suite S_i' after mutating becomes $<c_1$, c_2, **c_{13}, c_{12}**, c_{14}, c_{15}, c_{16}, c_8, $c_9>$.

4) The interacting testing among user sessions

Some new test cases, which contain the requested information of different users, can also be generated using crossover. The goal of the new test cases is to detect errors caused by the use of possible conflicting data shared by different users. The crossover way shows the idea of information interchange among different user sessions. Let S be a test suite, the steps of generating a test case using crossover are:

(1) for a test case c in S, randomly generate a number $g \in [0, 1]$, if a given crossover probability is not less than g, then

① copy the requests from r_1 to r_i in c to an interim test case, where i is a random number, $i \in [1, |c|]$; $|c|$ is the length of URL trace in c;

② select a test case d (different from c) in S randomly, and then search r_j in d reversely, which is the first one to have the same URL as r_i. If not found, then another test case (also denoted as d) is selected, until it is found or up to a given time;

③ if d is found, then all the requests after r_j in d are appended to the interim test case, which is just the new test case generated; otherwise go (2);

(2) repeat (1) until each test case in the test suite is processed.

We search the requests in test case d with an inverse search method, for the length of greatest common prefix of two test cases c and d in the same test suite is often greater than 1; when the corresponding URL trace from r_1 to r_i is the prefix of this greatest common prefix, we can not obtain different test cases if using the sequential search method. For example, given two URL traces $c=u_1u_2u_4u_3u_5u_6u_8u_7$ and $d=u_1u_2u_4u_3u_7u_4u_5u_9$, we randomly copy $u_1u_2u_4$ from c to an interim test case t firstly (i.e., t equals to $u_1u_2u_4$ temporarily). Here, $i=3$ and $r_i=u_4$. Now, we search r_j in d reversely, which is the first one to have the same URL as r_i and then we have $j=6$ and $r_j=u_4$. So, all the requests after r_6 in d (i.e., u_5u_9) are appended to t such that $t=u_1u_2u_4u_5u_9$. If searching the requests in d with a sequential search method, we have $j=3$ and $r_j=u_4$. This time, if copying all the requests after r_3 in d (i.e., $u_3u_7u_4u_5u_9$) to t, we have $t=u_1u_2u_4u_3u_7u_4u_5u_9$, which equals to d itself. And no new test case is generated. The reason is that the corresponding URL trace (i.e., $u_1u_2u_4$) from r_1 to r_i ($i=3$) in c is the prefix of greatest common prefix (i.e., $u_1u_2u_4u_3$) of c and d.

In addition, it is often not to mutate a test case, for the new generated test case after exchanging the two adjacent requests r_i and r_{i+1} makes no sense. The possible reason is that the former request r_{i-1} may never reach the request r_{i+1} (note that, this time r_i becomes the next request of r_{i+1}).

IV. EXPERIMENTAL ANALYSIS

Consider a typical miniature Web application that we have developed to demonstrate our approach: the SWLS (Simple Web Login System) shown in Fig. 4.

Starting at the first page (indicated by a dashed arrow, reasonably, a *blank* page can be used to request for the first page of a Web application), i.e., a *home page (p_1)*, the user can enter into the *news page (p_2)* to list the news by clicking on the *view* link, or enter into the *login page (p_3)* by pressing the *login* button. In page p_3, the user enters the *userid* and *password,* and presses the *submit* button. Upon this pressing, the *userid* and *password* are sent to the Web server for authentication. A *logged page (p_4)* will be loaded if both *userid* and *password* are correct. On the contrary, an *error page (p_7)* containing an error message is displayed if at least one of the submitted values for *userid* and *password* is wrong. From the *logged page*, it is possible to go to *info page (p_5)* for secure information viewing by just clicking on the *browse* link. The user can click on the intra-page link *continue* to view the different parts of the same page p_5 if it is too long. A *logout page (p_6)* will be displayed when

the user presses the *exit* or *logout* button. The user may come back to the *home page* for login again. Note that each time the *login page* is displayed, both the *userid* and *password* fields should be initialized to be empty.
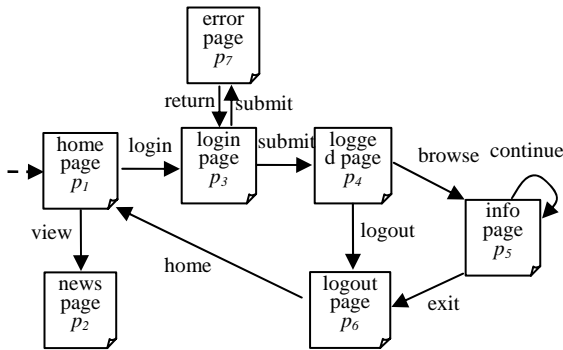

Figure 4. A simple Web login system

We inject only one fault in each page respectively, so at least 7 faults exist totally (there may be other faults in the Web application originally). A set of user sessions are obtained after scanning user logs on the Web server and purging their irrelevant information; then 89 meaningful user sessions are finally created after scanning them again. Only 17 user sessions[†] are used to generate test cases after the reduction of the meaningful user sessions using the URL trace-based ReduceUSession algorithm; and the reduction ratio reaches 80.9%. Impersonally, more user sessions there are for a given Web application, much higher is the reduction efficiency, for more user sessions mean higher possibility that the URL trace requested by a user session is the prefix of that requested by another in the view of statistics. The set of 17 user sessions (or test cases) is denoted by Γ1. After grouping and prioritizing Γ1, an initial executing sequence <S1, S2, S3> of test suites is obtained, which is denoted by Γ2, where S1, S2 and S3 contain initial executing sequences of 7, 6 and 4 test cases respectively. We achieve the final executing sequence <S1', S2', S3'> of test suites using genetic algorithm for further processing, which is denoted by Γ3. That's to say, S1', S2' and S3' is obtained through crossovering and mutating S1, S2 and S3.

Now, we run the test suites (and test cases) in Γ1, Γ2 and Γ3 for the Web application respectively, and find all the 7 faults injected as well as an additional fault (i.e., the user may browse p2 just by directly entering its address in the URL address bar). The executing time of Γ2 and Γ3, however, is much shorter than that of Γ1, and it takes less time of Γ3 than that of Γ2 too. This means that the test suites (and test cases) grouped and prioritized run faster than those, which are not grouped and prioritized; and that the test suites (and test cases) processed further by genetic algorithm are much faster. Table I contrasts some features of Γ1, Γ2 and Γ3. It is predictive that the test approach proposed will yield more evident positive effects for larger Web applications.

[†] For the convenience of test run, we have preprocessed the user sessions appropriately.

Table I. The feature contrast of Γ1, Γ2 and Γ3

	Γ1	Γ2	Γ3
reduced	yes	yes	yes
grouped and prioritized	no	yes	yes
optimized by genetic algorithm	no	no	yes
execution sequence	randomly	<S1, S2, S3>	<S1', S2', S3'>
execution time	long	short	shorter
fault detection ability	weak	strong	stronger

V. DISCUSSION AND EXPLANATION

The above experimental results show that the test cases generated by applying our algorithms do not increase the number of faults detected. It takes, however, less time to find the same faults by our approach. In this respect, the approach in this work is of higher fault detection ability than others. After obtaining a large volume of user sessions from Web servers, we eliminate *redundant* user sessions using the URL trace-based ReduceUSession algorithm and preserve *necessary* user sessions, to cut down the number of test cases. Then, we group and prioritize similar user sessions according to their common prefix, to achieve different initial test suites and test cases, and their initial executing sequences. The test scheme generated by the elementary prioritization is not much approximate to the best one. Therefore, genetic algorithm is employed to optimize the grouping and prioritization. Applying genetic algorithm to the optimization of the test cases reflects the random access property of user requests. So, this approach should improve the quality of test cases.

We can crossover the genes chain of two parent individuals selected with the probability of p_c using the TSCrossover algorithm to yield two new next generations and mutate each bits of the genes chain of new individuals with the probability of p_m using the TSMutation algorithm. The two new next generations crossovered are used to replace their parents; and the new individuals after mutating are used to replace those before mutating. The reasons are:

(1) the new individuals (test suites) crossovered and mutated are better; they can find errors in less time when being executed; and

(2) whether the crossover or mutation of the test suites do not change the test cases; only their executing sequences are changed. The executing sequences of test cases can, however, make an influence on whether or not we find errors earlier and what types of errors.

Therefore, the process of crossovering and mutating test suites optimizes the final coverage ratio of errors. The process also do not yield more test suites and test cases. However, it can increase the test executing efficiency. As for the interacting test of user sessions using crossover, although the number of test cases will increase, the new test cases can detect errors caused by the use of possible conflicting data shared by different users. That's to say, the types of errors detected are different, the coverage ratio of errors increases and the quality of test cases improves.

In testing, we choose to run the test suite and test case with higher priority and then those with lower priority. We can also run the test suites concurrently. Once some test case in some test suite is executed and test requirements are satisfied, the remainder test suites and test cases are not executed yet. This lessens the executing time greatly. Note that, there is an executing precedence relationship among different test suites and test cases in each test suite. The user sessions (test cases) are grouped, therefore, the testers can choose to run several test suites first to find some types of errors in greater probability and in less time. So, our approach is sound and flexible.

VI. RELATED WORK

Some methods and techniques were presented [12-17] to test Web applications. Cynthia, et al. [12] attempted to show the merits of traditional software testing as applied to the domain of Web accessibility, and to make the case for a shift of focus from post hoc validation to proactive testing as the preferred means of assuring accessibility for Web applications. Jukka [13] discussed common concepts, practices and tools that lie at the heart of Web application performance testing. Jiang, et al. [14] presented a quick testing model of Web performance based on testing flow. They also contributed a testing method of the testing model combining with LoadRunner testing tool. Zhu, et al [15] addressed an approach to discovering possible inconsistencies caused by interactions with Web browser buttons and the property of a Web page related to Web browser buttons. Zeng, et al. [16] proposed a new automation testing framework based on the concept of object feature set and dynamic searching policy. Umar, et al. [17] discussed the interoperability and integration testing of Web-based applications. Most of these methods are achieved through extending the testing methods for traditional software. Additionally, none of these methods yields test data according to user sessions in practice.

Elbaum, et al. [18] demonstrated the fault detection capabilities and cost-effectiveness of user session-based testing. They found that user session techniques uncover certain types of faults but not faults associated with rarely entered data. They showed that increasing the number of collected user sessions helps to improve the effectiveness, but it also increases test cost. Increment concept analysis [19] is used to analyze user sessions dynamically and minimize continuously the number of user sessions maintained. In ref. [20], the genetic algorithm and formal concept analysis are combined to trace the relationship between test data and corresponding test run.

Sthamer [21] analyzed deeply the test case optimization efficiency of different coding schemes and fitness functions of genetic algorithm for different structure-based software in his doctoral dissertation, as provides much valuable experience for other researchers. Some researchers also studied on test case generation techniques using genetic algorithm [22-24]. However, Most of them aimed for simplex optimization focusing on one-off optimization computation, while the testing is continuous and iterative. So, dynamically continuous optimization computation is more propitious to improve test performance. Jia, et al. [25] discussed the key problems of producing test data of covering designated paths using genetic algorithm, and introduced deeply the factors of influencing the genetic algorithm's efficiency through experimental results. Krishnamoorthi, et al. [26] proposed a new test case prioritization technique using genetic algorithm. It prioritizes subsequences of the original test suite so that the new suite, which is run within a time-constrained execution environment, will have a superior rate of fault detection when compared to rates of randomly prioritized test suites. These studies [18-26] use genetic algorithm to analyze test problems with the exception of refs. [18] and [19], but not focusing on Web application testing. In ref. [18], the authors discussed user session-based Web application testing, but not concerning deeply the optimization of test cases, while Godin, et al. [19] used increment concept analysis to analyze user sessions, not using genetic algorithm.

Different from theirs, the approach in this work generates and optimizes test cases for Web application testing based on user sessions using genetic algorithm. We begin to analyze user logs on Web server to obtain user sessions, and then conduct a deep investigation into Web application testing. Our approach bears a certain guiding value whether in theory or in practice.

VII. CONCLUDING REMARKS AND FUTURE WORK

Generating test cases of high quality is the premise of Web application testing. Our approach captures a series of user events, i.e., the sequences of URLs and name-value pairs in Web server (s). It then employs the reduction, grouping, prioritization and genetic algorithm to yield test cases and optimizes them. Compared to the methods of capturing user events in clients, our approach is very effective when a large volume of users exist, and it is a Web application testing method of high efficiency. The main contributions include:

(1) an approach to generating and optimizing test cases is proposed for Web application testing based on user sessions using genetic algorithm. The testing process relates more to functional testing.

(2) a URL trace-based reduction algorithm is designed. The user sessions acquired are lessened greatly using the algorithm. However, it covers all the URLs requested by the original set of user sessions and keeps the sequence of URL requests.

(3) an approach to grouping and prioritizing user sessions is presented, as can improve the efficiency of test run.

(4) an approach to generating new test cases is proposed using crossover. The new test cases generated include information requested by different users and can detect errors caused by the use of possible conflicting data shared by different users. That's to say, the crossover indicates the idea of information interchange among different user sessions, which helps to test the

interacting of user sessions. The interaction testing relates more to usability and performance testing.

(5) a strategy of test reuse and concurrence is provided, as can decrease the time of test run greatly as well as lessen test cost.

Some characteristics of Web applications can be analyzed using user session-based testing approaches. Fox example, if the connections included in most sessions identified are very little in a Web application. It means that the users may usually obtain their necessary data using less TCP connections and the outer objects requested are also limited, or it takes the users enough time to accomplish some task in some Web page.

Web application testing is much complex systems engineering. It is not easy to acquire an effective and practical test scheme. Our approach only evaluates a test case according to its test coverage ratio. However, many factors need to be considered, such as the running cost of each test case itself (e.g., CPU time), including the actual running time, loading time, time to save test state, and the influence of different test criteria on a test case. All these questions need to be answered in future research.

REFERENCES

[1] Sprenkle S, Sampath S and Gibson E, et al., "An Empirical Comparison of Test Suite Reduction Techniques for User-session-based Testing of Web Applications", Technical Report 2005-09, University of Delaware, 2005.

[2] Elbaum S, Karre S, and Rothermel G, "Improving Web Application Testing with User session Data", Proc. of the 25th Int'l Conf. on Software Engineering, Portland, Oregon, May 2003, pp. 49-59.

[3] Sampath S, Mihaylov V and Souter A, et al., "A Scalable Approach to User-session Based Testing of Web Applications through Concept Analysis", Proc. of the Automated Software Engineering Conference, Sept 2004.

[4] Yoo S, Harman M, "Using Hybrid Algorithm for Pareto Efficient Multi-objective Test Suite Minimisation", Journal of Systems and Software, 2010, 83 (4), pp. 689-701.

[5] Jose C B R, Mario A Z R and Francisco F de V, "Test Case Evaluation and Input Domain Reduction Strategies for the Evolutionary Testing of Object-Oriented Software", Information and Software Technology, 2009, 51 (11), pp. 1534-1548.

[6] Zhong H, Zhang L and Mei H, "An Experimental Study of Four Typical Test Suite Reduction Techniques", Information and Software Technology, 2008, 50 (6), pp. 534-546.

[7] Zhang L, Hou S S and Guo C, et al., "Time-aware Test Case Prioritization Using Integer Linear Programming", Proc. of the 18th Int'l Symp. on Software Testing and Analysis, 2009, pp. 213-224.

[8] Korel B, Koutsogiannakis G and Tahat L, "Application of System Models in Regression Test Suite Prioritization", Proc. of ICSM, 2008, pp. 247-256.

[9] Zhang X F, Xu B W and Nie C H, et al., "An Approach for Optimizing Test Suite Based on Testing Requirement Reduction", Journal of Software, Apr. 2007, 18 (4), pp. 821-831 (in Chinese with English abstract).

[10] Engstrom E, Runeson P and Skoglund M, "A Systematic Review on Regression Test Selection Techniques", Information and Software Technology, 2010, 52 (1), pp. 14-30.

[11] Holland J H, "Adaptation in Natural and Artificial System", University of Michigan Press, Michigan, USA, 1975.

[12] Cynthia C S, Mike B, "Application of Traditional Software Testing Methodologies to Web Accessibility", Proc. of the 2010 Int'l Cross Disciplinary Conf. on Web Accessibility, ACM, USA, 2010.

[13] Jukka P, "Web Application Performance Testing", Master's Thesis, Department of Information Technology, University of Turku, 2009.

[14] Jiang G Z, Jiang S J, "A Quick Testing Model of Web Performance Based on Testing Flow and its Application", Proc. of the 6th Conf. on Web Information Systems and Applications, 2009, pp. 57-61.

[15] Zhu B, Miao H K and Cai L Z, "Testing a Web Application Involving Web Browser Interaction", Proc. of the 10th ACIS Int'l Conf. on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, 2009, pp. 589-594.

[16] Zeng W D, Jiang N K and Zhou X B, "Design and Implementation of a Web Application Automation Testing Framework", Proc. of the 9th Int'l Conf. on Hybrid Intelligent Systems, 2009, pp. 316-318.

[17] Umar F, Usman A, "Testing Challenges in Web-based Applications with Respect to Interoperability and Integration", Master Thesis, Blekinge Institute of Technology, Ronneby, Sweden, 2009.

[18] Elbaum C, Rothermel G and Karre S, et al., "Leveraging User Session Data to Support Web Application Testing", IEEE Transaction on Software Engineering, May 2005.

[19] Godin R, Missaoui R and Alaoui H, "Incremental Concept Formation Algorithms Based on Galois (concept) Lattices", Computational Intelligence, 1995, 11(2), pp. 246-267.

[20] Khor S and Grogono P, "Using a Genetic Algorithm and Formal Concept Analysis to Generate Branch Coverage Test Data Automatically", Proc. of the Int'l Conf. on Automated Software Engineering, 2004, pp. 346-349.

[21] Sthamer H H, "The Automatic Generation of Software Test Data Using Genetic Algorithms", PhD Thesis, University of Glamorgan, Wales, UK, 1996.

[22] Ahmed M A, Hermadi I, "GA-based Multiple Paths Test Data Generator", Computers & Operations Research, 2008, 35 (10), pp. 3107-3124.

[23] Srivastava P R, Gupta P and Arrawatia Y, et al., "Use of Genetic Algorithm in Generation of Feasible Test Data", ACM SIGSOFT Software Engineering Notes, 2009, 34 (2), pp. 1-4.

[24] Mateo F, Sovilj D and Gadea R, "Approximate k-NN Delta Test Minimization Method Using Genetic Algorithms: Application to Time Series", Neurocomputing, 2010, 73 (10-12), pp. 2017-2029.

[25] Jia X X, Wu J and Jin M Z, et al., "Some Experiment Analysis of Using Generic Algorithm in Automatic Test Data Generation", Journal of Chinese Computer Systems, 2007, 28 (3), pp. 520-525 (in Chinese with English abstract).

[26] Krishnamoorthi R, S A Sahaaya Arul Mary, "Regression Test Suite Prioritization Using Genetic Algorithms", Int'l Journal of Hybrid Information Technology, 2009, 2 (3), pp. 35-52.

Qian Zhongsheng was born in 1977. He received his M.S. degree from Nanchang University of China in 2002, and Ph.D in Computer Science at Shanghai University of China in 2008. He has also been with the School of Information Technology at Jiangxi University of Finance & Economics of China since 2002. His current research interests include software engineering and Web application testing, etc.