Proposed Automated Testing Framework for Concurrent Software (Part B)

by

**Team M1**

Greig Cairns

Junaid Shakoor

Mohamad Amine Belabbes

Liam Hourston

William McDonald

Table of Contents

[1. Introduction to Proposed Framework 3](#_Toc477534602)

[1.1 Driver Module 3](#_Toc477534603)

[1.2 Tester Module 3](#_Toc477534604)

[1.3 Testing Types and Categories & Banking System Specimen 4](#_Toc477534605)

[1.3.1 Resource Locking Tests (Wrong and No Lock) 4](#_Toc477534606)

[1.3.2 Non-Atomic Operation Tests 4](#_Toc477534607)

[1.3.3 Notify Operation Tests 10](#_Toc477534608)

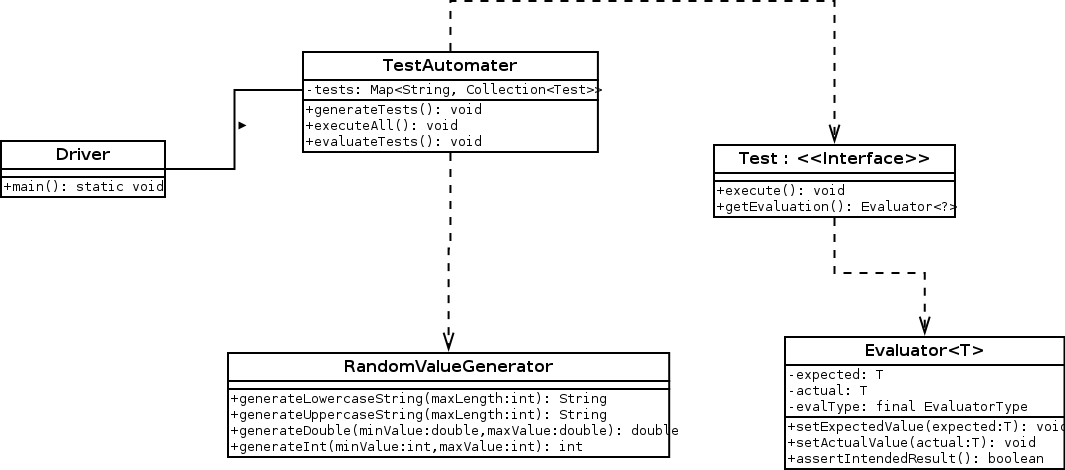
[1.3.4 Sleep Operation Tests 15](#_Toc477534609)

[1.4 Evaluator Module 3](#_Toc477534610)

[2. Results 4](#_Toc477534611)

# **Introduction to Proposed Framework**

The proposed framework, in theory, layouts the groundwork for a set of modules that work together to detect bug patters allowing developers to recognize and write better tests, validate, instrument. Following is a high-level view of the Proposed Framework.



The above diagram represents a high-level view of the proposed framework system for concurrent software testing. The parts being thoroughly implemented contain the tester module and the evaluator module. Together these two modules communicate with the driver module for system controlling. Following are some detailed description of each module, the types of bugs tested, and their final evaluation.

## 1.1 Driver Module

This module acts as the access point to the system and allows setting up of all resources. Essentially, this module is called Driver because it controls the setup, execution for all the modules in the proposed system, and retrieves the evaluation results after testing.

## 1.2 Tester Module

The proposed testing framework structure resembles as the figure above where different types of testing bugs implement the TEST interface which are being used by the TestGenerator and TestExecutor classes. This module, namely TestAutomater, is responsible for generating tests for all the test types we have in the Tests module. It then creates multiple instances to have a higher probability of catching bug patterns. This allows multiple category of bugs, as described in section 2.1 below, to be injected and tested with this structure.

The setup process also includes random generation of values which are passed as parameters to the test instances. The random generator class allows multiple types of values to be generated ranging from upper and lower case strings, ints, and doubles. This custom-made library’s methods require the range we want the value to come from and in other cases the maximum length of the string.

The tester module then goes ahead with the execution of ever generated test instance in a sequential loop. The final responsibility of the tester module is to generate the evaluated results that it obtains from the evaluation module which exists inside every test type’s instance and shows the result to the user.

## 1.4 Evaluator Module

The Evaluator Module has been constructed to be manually linked to the tests packages. The Evaluator has a generic state that just acts as a mediator for storing the expected and the actual results for comparison in a later stage of the processing. This module has a generic Type <T> that is defined at the time of its creation by every test class. One of the parameters the Evaluator expects for comparing is the custom defined *ASSERT\_EQUALS* and *ASSERT\_NOT\_EQUALS* enumerators. Based on the comparison it is requested to do by the user, it then evaluates if the test was passed or not.

The Evaluator then takes in all the testing that it has been asked to do throughout the system and then produces statistics based on tailor made specifics given by the user. At the end of execution, we can see based on the statistics on how many test samples ran and cross checked with the expected values.

# **Evaluation Results**

## 2.1 Testing Types and Categories

Following are the testing types and categories we injected into our system.

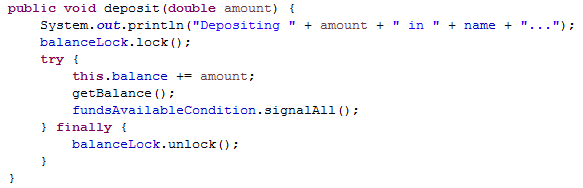
**Note**: The following code with examples may be subjected to slight changes and modification to work in accordance with the automated test generator module. Therefore, the example screenshots cited below may not contain all the information that is required for it to run as part of the system.

### 2.1.1 Resource Locking Tests

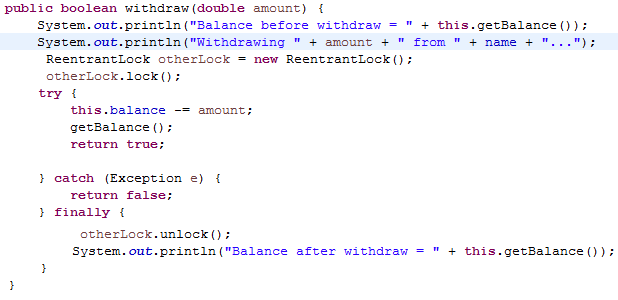
The wrong lock test is used to determine if two threads that operate at the same time are interfering with each other due to not properly locking certain shared resources as they both use a different type of lock that does not prevent the other from accessing it.

**TEST 1: Wrong Lock Pattern**

In our test for Wrong Lock pattern, the thread ‘dr’ is using the deposit() method on CurrentAccount ts1 inherited from account which uses the balanceLock reentrant lock as shown in figure 1.1 and thread ‘wr’ uses the withdraw() as defined in CurrentAccount method which uses the otherLock reentrant lock as seen in figure 1.1.1; they do not effectively lock the balance resource, meaning that there are occasions in operation where these locks can prevent the correct final balance from being calculated properly as the threads have acted in parallel rather than in series and one threaded lock condition has changed the balance while the other has been using it.



(*figure 1.1 A)*

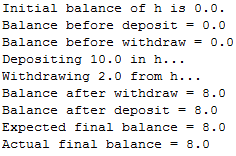


(*figure 1.1 B)*

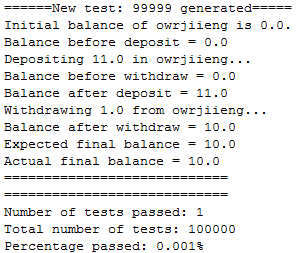
The evaluation that determines whether the bug has occurred or not is done by comparing the expected value to the actual value of the balance. The expected value is calculated as follows: 

Starting Balance + Money In - Money Out

This value is compared to the generated outcome of running two threads, one which uses deposit() on a CurrentAccount and one that uses withdraw(). If it is the same then the test passes, if not then something has gone wrong in the execution of the code, possibly the wrong lock bug.

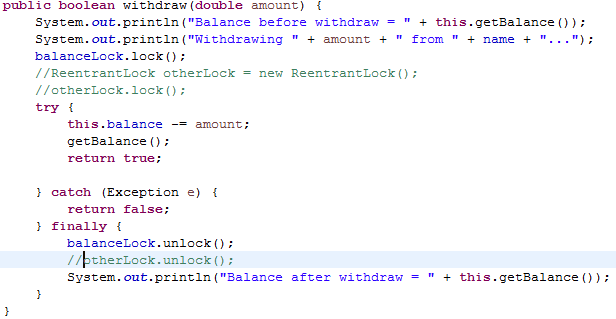
 (*figure 1.2)*

Output appears as shown in figure 1.2, this can be used to determine where a bug has occurred in operation if one does. In our testing of the code however we have found it hard to find occurrences of the wrong lock bug, after thousands of runs the expected and final balances were the same in all but a very few number of cases, which are impossible to locate through just the terminal but can be seen in figure 1.3.



(*figure 1.3)*

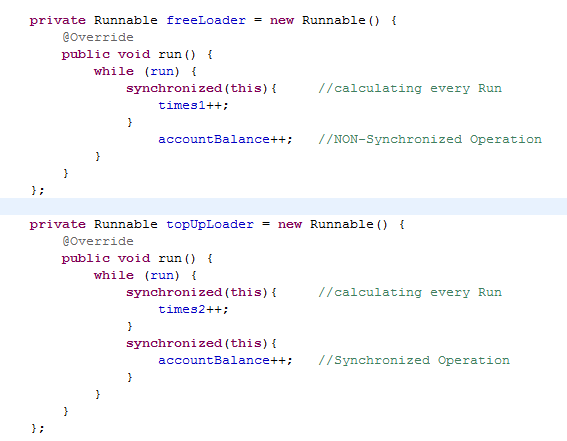
Any chance of this bug possibly occurring could be removed by altering the code of the withdraw method as seen in figure 1.4 so that it used balanceLock instead of otherLock.



(*figure 1.4)*

**TEST 2: No Lock Pattern**

In our test 2 for No Lock pattern, there’s a thread pool of a few number of threads consisting of two types: A thread that TopUpLoader increments the accountBalance with a synchronized block and the other thread FreeLoader that increments the accountBalance without a synchronized block. Both threads have a counter to check how many times they ran.

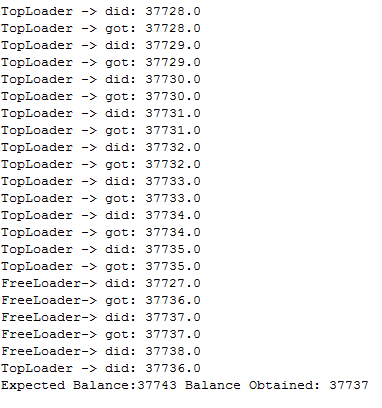


(figure 1.5)

At the end of the sample run we check that the counters from both those threads incrementing the starting accountBalance by one unit plus the starting balance itself would be our expected balance. 

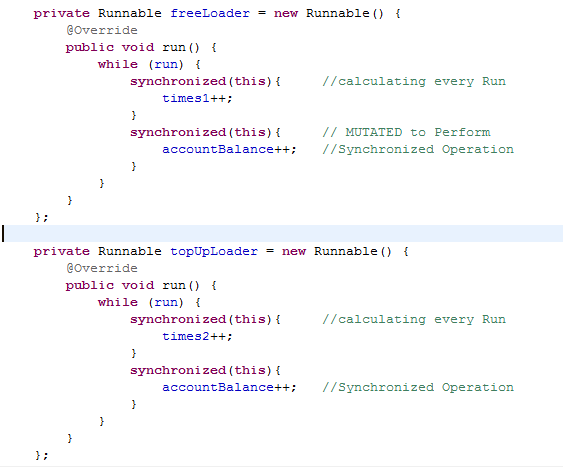
Starting AccountBalance + no. times TopUpLoader runs + no. times FreeLoader runs

Our expected account balance should be equal to the summation of these three fields and should not have any discrepancy between the two fields. If so, therefore we have encountered the no lock pattern. However, one of our test runs obtained the following results

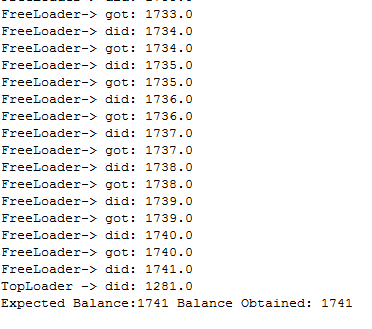


(figure 1.6)

However, fixing our bug pattern means we have to mutate the code into working according to how a standard, thread-safe, code structure should work. Therefore, we'll be surrounding all resources with a proper locking mechanism and ensure that the resources are locked and the operations perform when the resources become available. Following is an example of Runnable threads with mutated code and their respective output:



(figure 1.3)

 (figure 1.4)

### 2.1.2 Non-Atomic Operation Tests

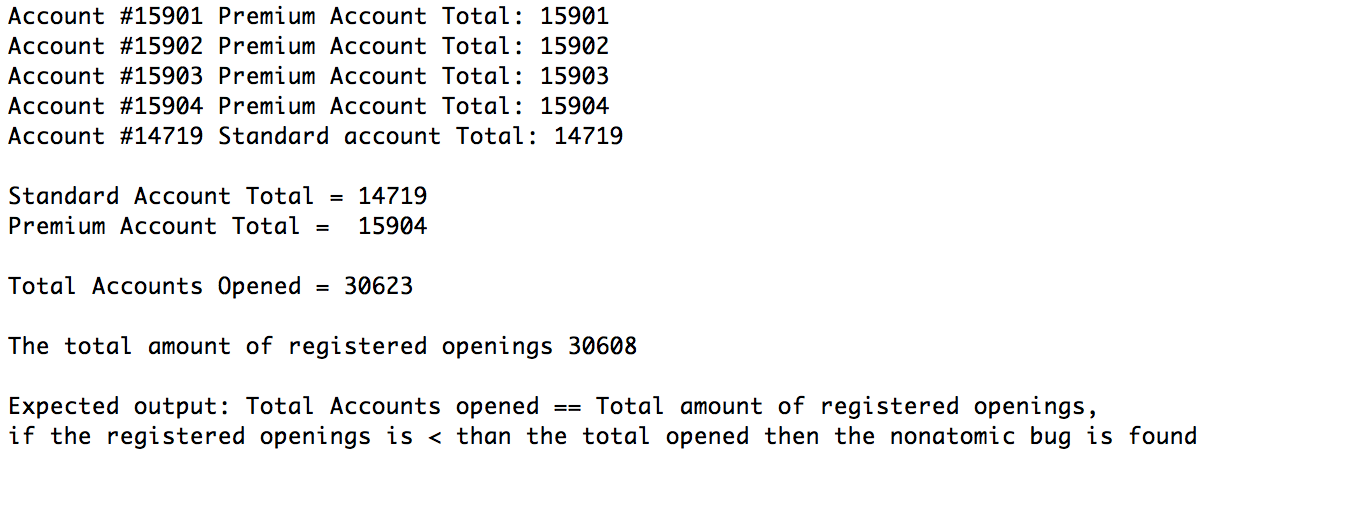
With this bug pattern injection, two different non-atomic operations are used to find instances of the bug. The approach with both operations is to create and execute two main threads and then use an executor service to run multiple 'worker threads', in a JAVA environment, to have a higher probability for the bug pattern to take place.

**Test 1:**

In our **test 1**for this category(Figure 2.2 – A, B) we’re using the x++ non-atomic operation as our example. The code keeps track of the amount of premium and standard accounts opened over a period of time. Two threads, one for each account type, will add to the total amount of each. when each thread is executed, a total amount of accounts opened is also executed. In theory, the total amount of premium/standard accounts should be equal to the total accounts opened but from the below test output this is not the case, proving the bug occurrence.



*(figure 2.2)*



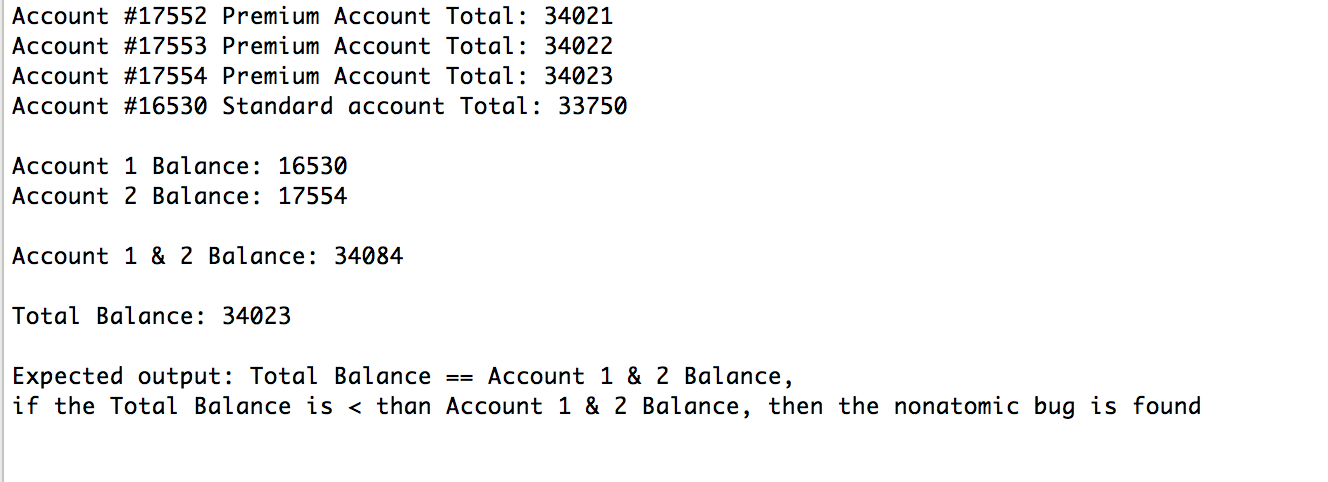
*(figure 2.3)*

In keeping with the theoretical use of a banking system, this would mean that in this test. A number of accounts either premium or standard were not registered opened. 

**Test 2:**

In our test 2 (figure 2.4, 2.5) we’re using the x+= non-atomic operation, the structure remains relatively the same. Two bank accounts that implement the += operator for adding to the account balance. The amount added to the balance each time the += operation is executed is a fixed amount. A total balance for the two accounts is executed withe same fixed amount each time either of the threads is executed.

 *(figure 2.4)*



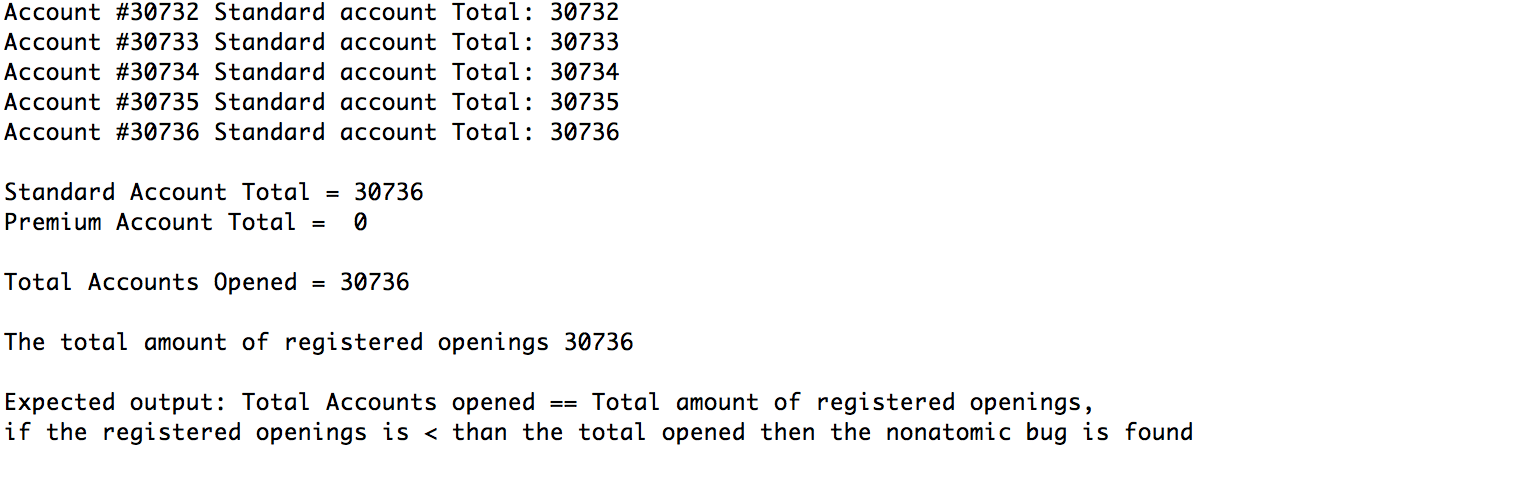
*(figure 2.5)*

In theory, the total balance should be equal to the balance of account1 + account2 but it is clear that this is not the case, once again proving the bug pattern occurrence.

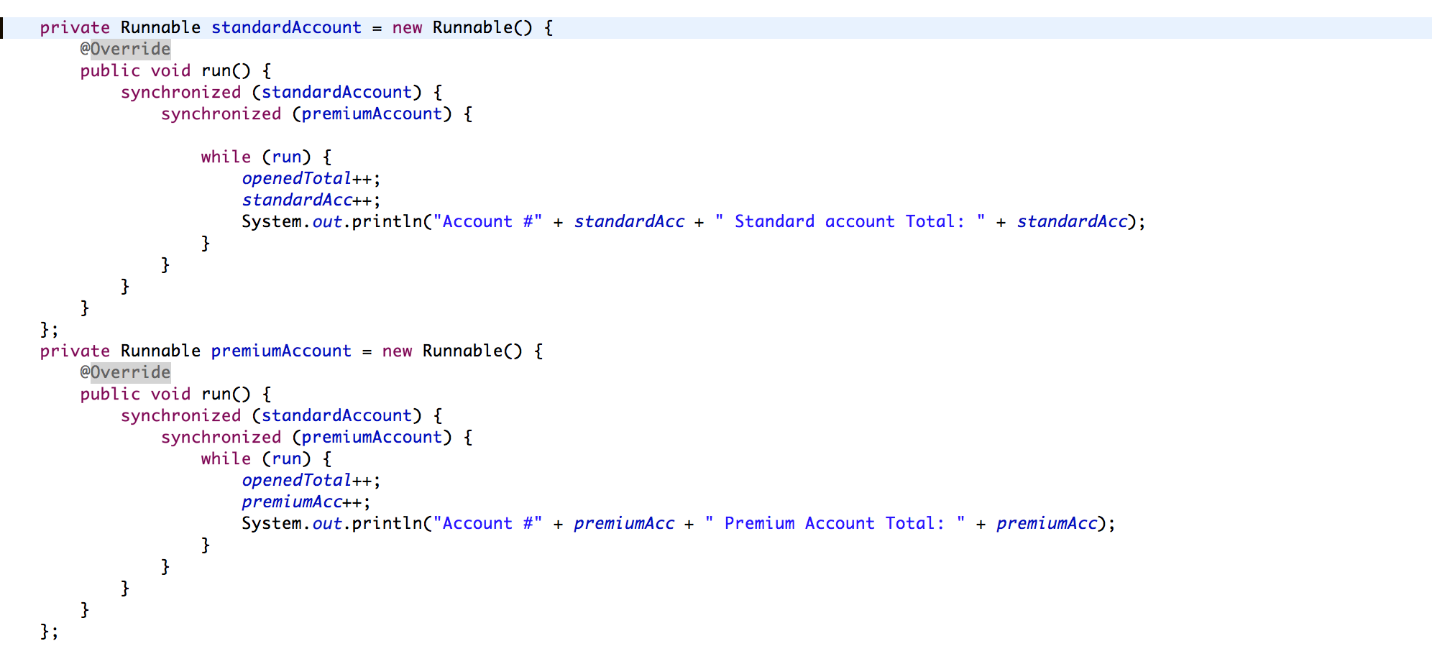
In keeping with the theoretical use of a banking system, this would mean that in this test. It would be possible to not only miss a deposit from a customer into their account, but could also mean that the total amount taken for the bank, for that period of time would also be wrong.

This bug can be resolved by the use of synchronization so that two or more threads are unable to access the same resource simultaneously. Meaning that each time the resources are updated, they are done correctly and none are updated at the same time resulting in only one update for that resource.

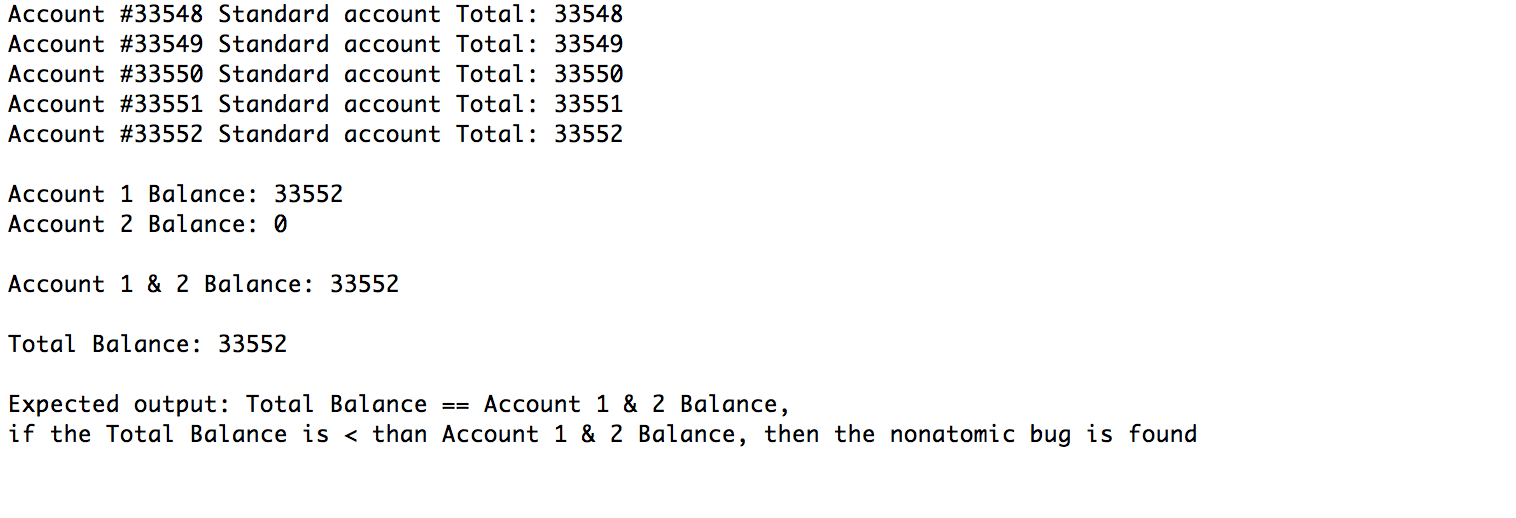
 *(figure 2.6)*



*(figure 2.7)*



*(figure 2.8)*

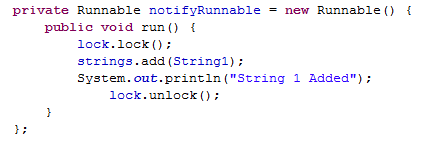


*(figure 2.9)*

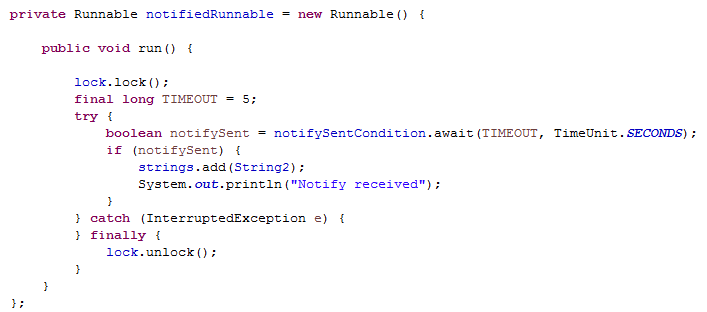
### 

### 2.1.3 Notify Operation Tests

In our test for a lost notify bug we created two runnables, *notifyRunnable (figure 3.1)*and *notifiedRunnable (figure 3.2)*, each one adds a string to the arraylist *strings*when run. These were each run by a separate thread, *notifiedRunnable*will only add it's string to *strings*if it is notified by *notifyRunnable*within 5 seconds.



*(figure 3.1)*



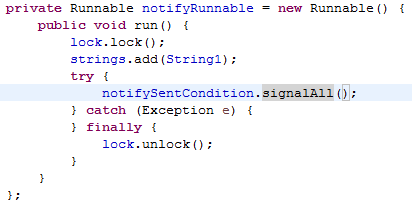
*(figure 3.2)*

The generated *string’s* array list is compared to another which was generated from the same strings to evaluate whether or not String2 has been added due to the notify being received.

C:\Users\Junaid\AppData\Local\Microsoft\Windows\INetCache\Content.Word\3.3.png

*(figure 3.3)*

Output can be seen in figure 3.3 showing a running where the notify was not received and string 2 was not correctly added to*strings* as a result*,*occurs on every running of the test. Without the notify being present then the test will pass every time, showing that the search for the lost notify bug has been successful. To prevent this the code could be altered as shown in figure 3.4.  While this does not entirely prevent the lost notify bug it does allow for some of the notifies to be picked up by the *notifyRunnable*and for the bug not to occur on some runs, we cannot remove it entirely due to the interleaving.



*(figure 3.4)*

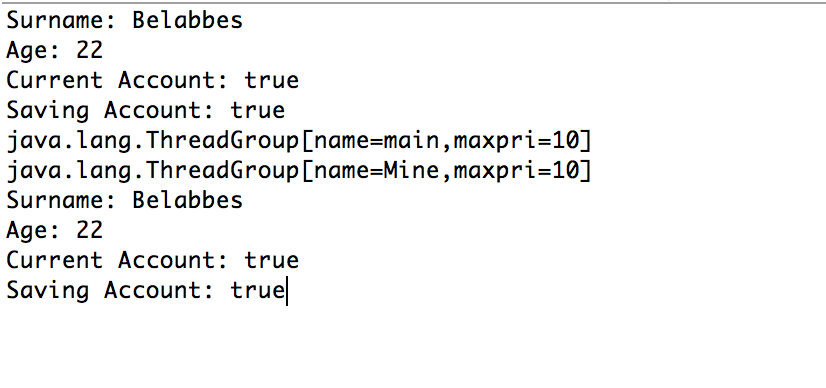
### 2.1.4 Sleep Operation Tests

Using methods such as sleep() or yield() forces the change of order of concurrent events and therefore, increase the probability of finding concurrent bugs.

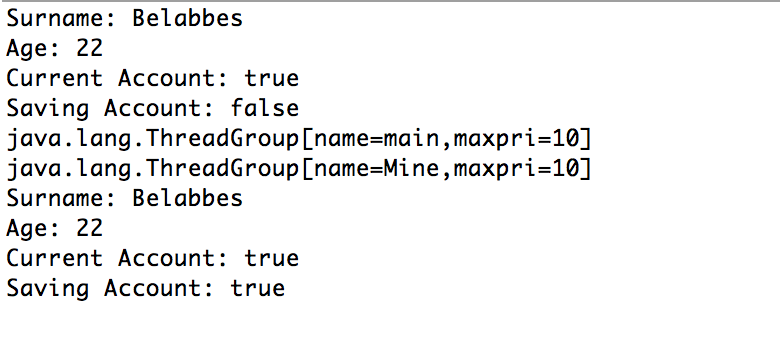
In the following example, the method move on is supposed to update the saving account to true when the age of the customer becomes greater than 20. We first start by creating threadGroup '*g1'*and '*g2'*. The threadGroup '*g1'* uses the main threadGroup as its parents while '*g2'*using '*g1'*as its parent. We create an object of type Bank called *'med'* and pass three arguments to the constructor which are: A string for surname, a Boolean for the current account and the age. Every object of type Bank has the Boolean value for the saving account initialized as false as standard.

We use two methods from the class Bank to illustrate the sleep() bug pattern: moveOn() which checks if the age is greater than 20; if so, the *Saving Account* becomes false. The second method is supposed to update the fields surname, current account (Boolean) value and age.

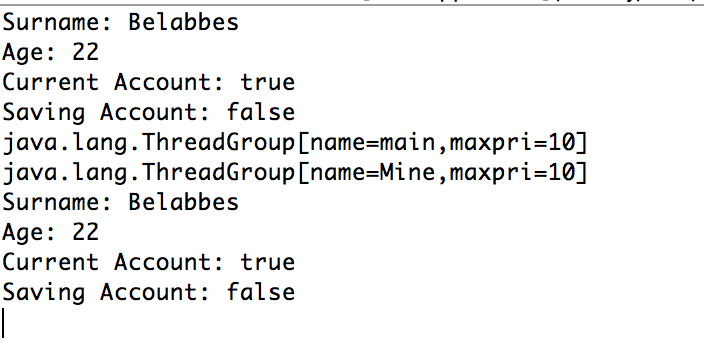
Furthermore, we place moveOn() in the parent group and update in the child group. First, we would like to update and then call moveOn() to check the new value of the age. So, we add a call to the thread to sleep a certain amount of time (0.10 seconds) just enough to let the child thread update the fields. We have two printing statements supposed to print the details of the customer, one before printing the thread details and one after. Both should show the same output when there is no bug:

 (*figure 4.1)*

However, some interleaving can happen when the bug is introduced such as:

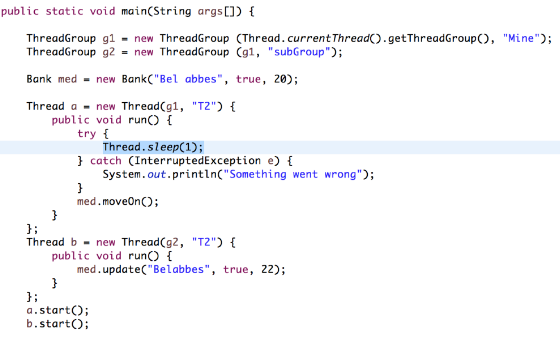


(*figure 4.2)*



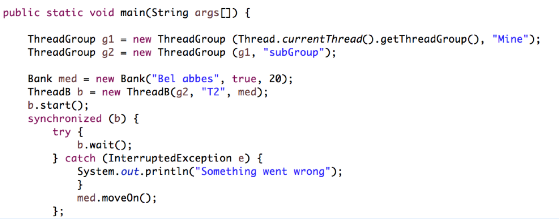
(*figure 4.3)*

Therefore, we can assume that we have encountered the bug pattern due to the sleep method calling in a concurrent environment.

*(figure 4.4)*

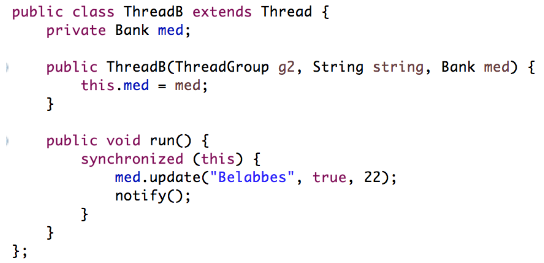
However, such bug can be resolved by using wait and notify. In the parent thread, we can introduce a synchronization mechanism and wait for the child thread until it finishes and notify the parent thread to take action. Here, the parent is thread a and thread b is the child thread, we use b.wait to wait for b to finish before taking action.

You



*(figure 4.5.1)*

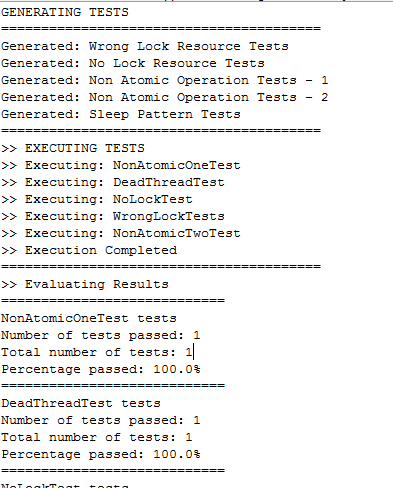
Meanwhile, thread b, we execute our operation update and call notify() so the parent thread a executes its method moveOn();

**

*(figure 4.5.2)*

## 2.2 Results

Following our detailed analysis of injection and mutation examples, we run all of our tests through the main tester module that has generated and executed all, and now the system pulls all evaluation results to show us all the tests that were executed and their pass and fail statistics. The system keeps an active count on the number of tests ran and their success rate as well as their percentage pass criteria. The following image shows us a glimpse of what to expect while the program is running.



*(figure 4.6)*

This allows the user to have an easier to understand high-level output of the stats without having to go into much detail. The evaluation module is constructed as such to accommodate command line as well as a graphical interface if built and would work seamlessly.