***ILP Testing Evidence document***

# **Testing techniques used**

Testing was performed at all 3 possible levels: ***unit***, ***integration,*** and ***system***. Examples of unit testing include test classes such as *TestOrder*, *TestRestaurant*, and so on; examples of integration testing include test class such as *OrderIntegrationTest*; the system test is in a class named *SystemTest*.

In terms of techniques being used, we can see the application of ***systematic functional testing*** in both the unit and integration test classes. The unit and integration tests were generated through ***specification-based testing***, which allows us to verify the functional requirements of the project related to the individual components, as well as integrating the real data from the REST server and checking that correct functionality is preserved.

Aside from this, we also see the use of ***structural testing*** techniques, such as ***equivalence partitioning***. This is clearly visible in the testing for order validation, since we need to account for several invalid labels (as well as a valid label) in order to ensure that the validation is working correctly, and the orders are sorted into the correct category according to the individual order details.

Another technique being used is that of ***boundary value analysis***. One example of where this can be seen is when testing edge cases for credit card details in order to ensure that the validation works as expected. Another example would be that of central area and no-fly-zones: checking edge case conditions, such as being exactly on the boundary for one of those areas, and ensuring the system makes the appropriate decision based on the requirements spec.

In order to assess how the testing measures the ***measurable attributes***, we will consider some of these attributes and how they were tested. For ***performance***, a timer was used to measure the system computation time and ensure that this falls within the required timeframe of 60 seconds. For ***functionality***, we considered all the requirements in the requirements specification, and how many of there were met after running the test suite. For ***fairness***, we considered all the orders delivered for a given date, and checked how many were delivered for each restaurant, in order to ensure that each restaurant had at least one order delivered.

# **Evaluation of adequacy of testing**

The testing that was performed is mainly ***optimistic***. For instance, if we consider the file writing (JSON and GeoJSON), these are tested at system level, but due to the large size of the files, we only perform automated checks to see if the files were created and do a manual check to see if the contents of the file appear to be correct. However, a malfunction of some sort on the REST server, the order validation mechanism, or the flight path planning algorithm could result in the contents of the files being incorrect/ faulty/ buggy. Another example of the testing being optimistic is in the case of the flight path planning algorithm: this is only checked during system testing, since we need real data to test the algorithm, and generating synthetic data would be too costly both timewise and resource-wise. Therefore, there could be missing cases that may not be considered and could have an impact on the reported level of functionality.

One evaluation criterion that can be used to motivate the choice of testing techniques is that of ***class coverage***. This criterion refers to the test suite accounting for a proportion of the classes in the system. This criterion justifies the testing technique of ***systematic functional testing*** since this technique enables us to test all individual classes/ components of the system as part of unit and integration testing.

Another criterion that can be used to motivate the choice of testing techniques is that of ***statement coverage***. This criterion refers to test suite executing a proportion of the executable statements in the program. This criterion justifies the testing techniques of ***equivalence partitioning*** and ***boundary value analysis***, since these techniques enable us to test the system more thoroughly, by considering several test cases.

Yet another criterion that can be used is that of ***test coverage***. This criterion refers to the test suite covering a good proportion of a proportion of the code for the system (classes and executable statements). Since it accounts for both aspects previously mentioned, this criterion justifies all the testing techniques used.

The last criterion that can be used to motivate the choice of testing techniques is that of ***specs coverage***. This criterion refers to the test suite accounting for a proportion of the requirements in the requirements specification. This criterion justifies the testing technique of ***systematic functional testing*** since this technique enables us to perform ***specification-based testing*** for the system’s functional requirements.

# **Results of testing**

One of the main issues encountered whilst testing the mentioned requirements was to do with ***order validation***. The issue here was that too many orders were labelled as invalid, and therefore the program was unable to work as expected. For instance, since too few orders were labelled as valid, sometimes all orders for a give restaurant would get lost; this had a huge impact on ***fairness***. This was resolved by creating an automated piece of code that looped over all orders on the REST server and printed out how many orders fell into each of the 10 possible categories. In this way, I was able to identify the source of the issue, which was my credit card validation, and fix the code accordingly.

Another issue detected whilst testing was related to the ***drone avoiding no-fly-zones***. This issue essentially consisted of the fact that the drone disregarded the no-fly-zones and went directly through them when making deliveries, which defied the whole purpose of the system and clearly impacted ***functionality***. This was resolved through debugging, with the source of the problem being that the no-fly-zones were never initialized, and so the drone had no way of realizing that no-fly-zones actually existed.

However, a follow-up issue to the previous issue was that, once the no-fly-zones were initialized, the runtime of the system was above the required “60 seconds or less”. This, of course, was related to the ***efficiency*** requirement, and severely impacted ***performance***. The actual issue here was that the program was stuck in an infinite loop whenever the drone got really close to a no-fly-zone. Due to the nature of the algorithm (a greedy algorithm) the drone would constantly fluctuate between the two same points on the map whenever it got close to one edge of one of the no-fly-zones. This issue was fixed by adding an additional if-statement to check if the position had already been visited, which ensured the system would never get stuck in an infinite loop, and also solved the runtime issue.

# **Evaluation of testing results**

The ***class coverage*** criterion ensures that a good proportion of the classes in the system have been accounted for in the test suite. This means that the design of the system has been tested, and all components appear to function accordingly. Since all individual components function accordingly, this reduces the possibility of bugs/ faults arising when running the code, giving us more confidence that the testing is adequate, and the system works as expected.

The ***test coverage*** and ***statement coverage*** criteria ensure that a good proportion of the actual code is being executed and therefore tested. This means that the multiple scenarios/ cases are being accounted for and tested, and the behaviour of the system in these scenarios is noted and compared with the expected behaviour. In doing so, we can detect any faults in the code more easily, and therefore also correct these faults. By having a good test coverage, we can be sure that most possible outcomes are being considered, and, assuming the system works as expected in all scenarios, we can be more confident in our testing.

The ***specs coverage*** criterion ensures that a good proportion of the requirements in the requirements specification have been accounted for in the test suite. This means that we can account for how many functional requirements are covered by our test suite, and ensure that, for all those requirements, the system works as expected. In doing so (of course, assuming a high specs coverage), we can have more confidence that our system complies with the specification and therefore works as it should.

However, one major drawback of all these coverage measures is that they only server as a proxy for the thoroughness/ depth of the test suite. This is because it is very difficult to measure how adequate a test suite is for a given system, and so using these proxies, though not perfect, can give us measurable estimates and a good degree of confidence that the system works as expected.