

Metamorphic Testing in Autonomous System Simulations

Jubril Gbolahan Adigun
Department of Computer Science
University of Innsbruck
Innsbruck, Austria
jubril.adigun@uibk.ac.at

Linus Eisele
Department of Computer Science
University of Innsbruck
Innsbruck, Austria
eisele.linus@gmail.com

Michael Felderer
Department of Computer Science
University of Innsbruck
Innsbruck, Austria
michael.felderer@uibk.ac.at

Abstract—Metamorphic testing has proven to be effective for test case generation and fault detection in many domains. It is a software testing strategy that uses certain relations between input-output pairs of a program, referred to as metamorphic relations. This approach is relevant in the autonomous systems domain since it helps in cases where the outcome of a given test input may be difficult to determine. In this paper therefore, we provide an overview of metamorphic testing as well as an implementation in the autonomous systems domain. We implement an obstacle detection and avoidance task in autonomous drones utilising the GNC API alongside a simulation in Gazebo. Particularly, we describe properties and best practices that are crucial for the development of effective metamorphic relations. We also demonstrate two metamorphic relations for metamorphic testing of single and more than one drones, respectively. Our relations reveal several properties and some weak spots of both the implementation and the avoidance algorithm in the light of metamorphic testing. The results indicate that metamorphic testing has great potential in the autonomous systems domain and should be considered for quality assurance in this field.

Index Terms—metamorphic testing, simulation, cyber-physical system, test case generation, autonomous system

I. INTRODUCTION

There is a growing adoption of autonomous systems (ASs) in our daily lives and this requires that software and system engineers ensure that ASs are adequately tested. One of the now frequently used methods for testing is simulation testing in which different approaches such as metamorphic testing (MT), fuzz testing, search-based testing, machine learning testing and model-based testing are conducted. In this work, we focus on the metamorphic testing technique as it has effectively been used across many domains [1], [2]. Evidently, MT has been receiving an increasing amount of attention and has led to several successful implementations, but not as many applications in autonomous systems. Moreover, there appears to be a limited amount of studies when it comes to the identification of metamorphic relations (MRs) despite there being several open questions on how to construct effective MRs [3]. We believe that the lack of applications in autonomous systems is unjustified since MT may be a very appropriate strategy for

testing autonomous systems. Consequently, the contributions of this work are;

- We introduce MT in autonomous systems and define several MRs.
- We develop an obstacle avoidance task in an autonomous drone simulation.
- We implement the test cases to demonstrate the defined MRs.
- We provide detailed technical information for setting up the simulation used in this work.
- We provide evidence that MT is an adequate testing strategy in the simulation of ASs

The remaining part of this paper is organised as follows: Section II presents a theoretical background to some of the concepts mentioned in this study. The approach in terms of the testing technique is described in Section III. In Section IV, we describe the implementation infrastructure as well as the tools used for the simulation and discuss the results in Section V. Further, we present some related work in metamorphic testing and highlight studies carried out using simulations in Section VI. Finally, we conclude and present some ideas for future work in Section VII.

II. BACKGROUND

Here, we provide background information about the concepts used throughout the paper.

A. Software Testing

Software testing is an essential practice widely applied during development in order to detect faults in a program. This usually consists of running the program under test on selected input data, producing corresponding output data which are then examined. The next step is to determine whether the program is correct or not. The mechanism which reveals this is known as the (test) oracle, and the common assumption that an oracle is trivially given is called the oracle assumption [4]. While this assumption is justified in many cases, there do exist programs that produce complex output such as compilers, simulators, or search engines. For these programs there may not exist an oracle or it may not be feasible to use, which we refer to as the oracle problem [2]. Programs which suffer from the oracle problem are often referred to as “untestable” since one

This work was partially supported by the Austrian Science Fund (FWF), under grant I 4701-N

can hardly benefit from testing when it is impossible or too inefficient to classify a program's output as (in)correct [4]. Autonomous systems are a critical domain in this regard since it can be immensely difficult to decide whether the behaviour of an autonomous system meets its specified requirements. Furthermore, autonomous drones are becoming increasingly popular for purposes like surveillance or search operations, while software testing in this field is complex [5].

B. Simulation Testing

Simulation testing allows designers, developers, engineers, and researchers to mimic real world systems and operational behaviours in a controlled virtual or hybrid environment. It has gained popularity in the scientific world and been applied severally in various complex cyber-physical systems (CPSs) such as in the autonomous vehicles [6], Internet of Things (IoT) and distributed systems domains. Contemporary testing methods such as rule-based techniques, semi-formal and formal verification often require trivialising systems and do not always provide adequate reliability necessary when testing complex systems [7]. More so, these models operate at a rather abstract level depicting states of the system or some other timed automata without giving insight about the operational environment in which a system would act. In this case, simulations offer a more robust exploratory platform as physical effects and environmental details can be observed and analysed. They can be used for understanding functional requirements of a system as well as to analyse the safety concerns of the system in operation as done by Huck et al. [7]. Simulations immensely reduce cost, time and risk for development and testing compared to a real-world environment and generally, an untested implementation should not be blindly applied to a real AS.

C. Metamorphic Testing

Metamorphic Testing (MT) is a testing technique that aims to alleviate the oracle problem by checking whether the inputs and outputs of a program across multiple executions satisfy certain relations, rather than directly investigating its input-output behaviour, as conventional testing strategies would do [2]. These relations, basically properties of the system under test, are known as Metamorphic Relations (MRs) [1]. For instance, suppose one were testing an e-commerce site and performs an initial search for a particular product across all regions where the business is located. It is not feasible to verify if the search result is fully correct, but one could use a follow up action to determine if the search result is problematic or not. So, one could note the set of items returned, say X , from the initial search. One may then apply another search filter e.g. a specific location filter to the search parameters such as "Region: My Location" and then run the search again. we expect that the returned set of products, say Y should be a subset of the former result X i.e. $Y \subset X$. This represents a simple example of a metamorphic relation where we may tell that something is wrong in our system if for instance, the cardinality of Y were greater than the cardinality of X even

though we may not know exactly where the error may have come from.

III. PRELIMINARIES

In this section, we define the relevant terminologies used in MT. We also describe the process of identification of metamorphic relations and (MRs) and briefly discuss how to construct effective MRs.

A. Terminology

We combine the notations used by Chen et al. in [8] and [9]. The resulting notation has been slightly generalised and further elaborated in some aspects.

Let $i, n, m \in \mathbb{N} \wedge m > n > i$, and let P be a program that implements an algorithm or function f . Given x_i as input for f , the respective output is denoted as $f(x_i)$. The result of executing P with test case \bar{x}_i is denoted by $P(\bar{x}_i)$. For a test case \bar{x}_i with the result $P(\bar{x}_i)$, a number of "next" test cases denoted as $\bar{x}_{i+1}, \bar{x}_{i+2}, \dots, \bar{x}_n$ may be constructed. A sequence of elements x, y, z (i.e. test cases, inputs, outputs, or results) is denoted as $\langle x, y, z \rangle$.

Definition 1 (Metamorphic Relation).

A metamorphic relation (MR) is a necessary property¹ of f over a sequence of inputs $\langle x_i, x_{i+1}, \dots, x_n \rangle$ and their corresponding outputs $\langle f(x_i), f(x_{i+1}), \dots, f(x_n) \rangle$. It can also be expressed as a relation R , where $R(x_i, x_{i+1}, \dots, x_n, f(x_i), f(x_{i+1}), \dots, f(x_n)) \Rightarrow \langle x_i, x_{i+1}, \dots, x_n, f(x_i), f(x_{i+1}), \dots, f(x_n) \rangle \in R$.

Definition 2 (Source Test Case and Follow-Up Test Case).

Consider a given MR R with $R(x_i, x_{i+1}, \dots, x_n, f(x_i), f(x_{i+1}), \dots, f(x_n))$. All inputs x_i and outputs $f(x_i)$ for $0 \leq i < n$ are referred to as *source inputs* and *source outputs*, respectively, and the correlating test cases \bar{x}_i as *source test cases*. From there, suppose that new test cases \bar{x}_j are created based on the existing source test cases, with $m > j > n$. Then all x_j are referred to as *follow-up inputs* and their test cases \bar{x}_j as *follow-up test cases*. A sequence of source and follow-up test cases (or inputs) is called a *Metamorphic Group (MG)*, i.e. $\langle \bar{x}_i, \bar{x}_{i+1} \rangle$ or $\langle x_i, x_{i+1} \rangle$.

Definition 3 (Metamorphic Testing).

Let R be an already identified MR with $R(x_i, x_{i+1}, \dots, x_n, f(x_i), f(x_{i+1}), \dots, f(x_n))$. Metamorphic testing (MT) of P based on R involves the following (remaining²) steps: [9]

- 1) Project the found MR onto the implementation, i.e. replace f by P in R and use the inputs as test cases.
- 2) Execute P applying the source test cases $\langle \bar{x}_i, \bar{x}_{i+1}, \dots, \bar{x}_n \rangle$ in order to obtain their respective outputs $\langle P(\bar{x}_i), P(\bar{x}_{i+1}), \dots, P(\bar{x}_n) \rangle$.
- 3) Construct and execute a number of follow-up test cases $\langle \bar{x}_j, \bar{x}_{j+1}, \dots, \bar{x}_m \rangle$ in accordance with the relation R and obtain their respective outputs $\langle P(\bar{x}_j), P(\bar{x}_{j+1}), \dots, P(\bar{x}_m) \rangle$.

¹A property that can be logically deduced from the algorithm [9].

²MR identification is also a step within MT, to be precise [8].

- 4) Decide whether R is violated. If R is not satisfied, then R revealed that P contains faults. Formally,
- $$\langle \langle \overline{x_i}, \overline{x_{i+1}}, \dots, \overline{x_n}, P(\overline{x_i}), P(\overline{x_{i+1}}), \dots, P(\overline{x_n}) \rangle, \langle \overline{x_j}, \overline{x_{j+1}}, \dots, \overline{x_m}, P(\overline{x_j}), P(\overline{x_{j+1}}), \dots, P(\overline{x_m}) \rangle \rangle \notin R \Rightarrow P \text{ is faulty}$$

B. Metamorphic Relation Identification

Previous studies found that due to the pitfalls involved in MR identification, the process constitutes a major challenge in MT [9]. One commonly misunderstood concept is that, albeit MRs being necessary properties of the target program, not all necessary properties are MRs. Another misconception is that not all MRs are equality relations, although many of them are vastly defined as such. An invaluable skill within the process of MR identification is the ability to distinguish good MRs from less effective ones, as deducible from several studies [1], [10], [11]. For instance, Chen et al. [11] emphasised that MRs should make the execution of follow-up test cases as different as possible from that of the source test cases, which has been confirmed by several later studies. They also reported that knowledge of the problem domain and the program structure is required for the determination of good MRs, whereas purely theoretical knowledge of the problem does not suffice. The survey further concludes that MRs ought to be formally described, as for instance pointed out by Hui and Huang [12]. Another important property of MRs is diversity (i.e. they should involve distinct parameters) as suggested by Liu et al. [13]. Mayer and Guderlei [10] have also conducted a study on the selection of good MRs and define good MRs to be rich in semantics of the system under test, and that being too close to the implementation limits the usefulness of MRs (e.g. when the difference between the source and follow-up test cases is balanced out by the implementation) [1], [10].

In the following we present two of the MRs that we defined for this project. Further, we add the following notation for specific parameters concerning an output $f(x_i)$:

$f(x_i)_d$ the distance travelled by the drone(s) in metres,
 $f(x_i)_t$ the wall clock time passed in seconds,
 $f(x_i)_{ac}$ the number of performed avoidance manoeuvres,
 Δ_d, Δ_t tolerance values ≥ 0 for distance and wall clock time;

1) Source Inputs and Follow-up Inputs:

Source Input	Follow-up Input
x_1 : obstacles, one drone, path A \rightarrow B	x_2 : obstacles, one drone, path B \rightarrow A
x_3 : no obstacles, two drones, separate paths $A_1 \rightarrow B_1$, $A_2 \rightarrow B_2$	x_4 : no obstacles, two drones, separate paths $B_1 \rightarrow A_1$, $B_2 \rightarrow A_2$

2) Metamorphic Relations:

$R_1 : \langle x_1, x_2, f(x_1), f(x_2) \rangle \in R_1 \iff f(x_1)_{ac} \geq 1 \wedge f(x_2)_{ac} \geq 1.$

This relation is aimed at testing the overall reliability of object avoidance using stationary obstacles. The number of required avoidance manoeuvres might differ due to an asymmetrical shape of the obstacle and/or the locations of A and B. The

constraint that the number of avoidance manoeuvres should be at least one for each path has therefore been chosen and shall ensure that avoidance is active and working.

$R_2 : \langle x_3, x_4, f(x_3), f(x_4) \rangle \in R_2 \iff$ for each drone: $f(x_3)_{ac} = f(x_4)_{ac} = 0 \wedge f(x_3)_d = f(x_4)_d \pm \Delta_d \wedge f(x_3)_t = f(x_4)_t \pm \Delta_t.$

This is one of the relations that involves two drones simultaneously. The object influence threshold and drone positions are chosen in a way such that the drones are separated just widely enough not to trigger object avoidance. Hence, we demand that the avoidance manoeuvre count should stay zero as object avoidance would soon be triggered once a drone wanders from its shortest path. Moreover, the travelled distances and taken durations should also be approximately identical for each drone among both paths.

IV. IMPLEMENTATION

In this section, we describe the system architecture, tools and frameworks, and APIs used as well as the parameters used within the simulator to run our simulation. Since our aim is to investigate the applicability of metamorphic testing in autonomous systems using simulations, we demonstrate the testing approach by implementing an object avoidance task, which is a non-trivial exercise in ASs.

A. System Setup

The technical setup consists of the simulation environment with the autonomous drone and required plugins, the ArduPilot software along with MAVProxy and the software in the loop (SITL) simulation, and the robot operating system (ROS) with the MAVROS package, using the MAVLink (Micro Aerial Vehicle Link) communication protocol [14]. For programmatically controlling the drone as well as performing obstacle detection and avoidance, the DroneKit [15] and GNC (Guidance Navigation and Control) [16] APIs are used. The entire simulation and implementation has been conducted on a local system running Ubuntu 20.04.2 LTS 64-bit.

1) *Gazebo*: One of the most popular simulators in robotics which was used to implement the simulation environment for this project. In addition to its popularity, we considered its superior online documentation and community support in our decision to opt for Gazebo [17]. Common alternatives are Unity [18], Webots, Simbad, or MRDS (Microsoft Robotics Developer Studio) [19]. We use given open-source models from Open Robotics³. Green landing platforms are used as starting points, while red landing platforms act as destination points.

2) *ArduPilot SITL*: ArduPilot is an open-source autopilot framework which supports a wide variety of hardware such as various aircraft, ground vehicles, or underwater vehicles [20]. A popular alternative is PX4 autopilot, while ArduPilot subjectively seems to be tested more thoroughly and to be more reliable [21]. The ArduPilot SITL (Software In The Loop) simulation provides a virtual way of using the vehicles,

³Available online at <http://models.gazebosim.org> [accessed Jan-2022].

allowing developers to test their behaviour realistically without hardware.

3) *MAVProxy*: A minimalistic ground control station for MAVLink⁴-based systems. It allows the operator to control the drone through setting parameters or directly sending commands. It is predominantly used by developers as it is command-line based and not very intuitive to use, however it can be complemented with another ground station like QGroundControl or Mission Planner, which come with a GUI (Graphical User Interface). The MAVProxy console is launched along with SITL which is able to read the messages sent from MAVProxy.

4) *Robot Operating System (ROS)*: An open-source robot operating system proposed by Quigley et al. in 2009 [22] which offers a communication layer above the host OS. It comes with a set of very useful tools for developing and programming robots. For this project, ROS Noetic was used along with the Catkin build environment for ROS packages. Essential ROS concepts are nodes, messages, topics, and packages (e.g. MAVROS).

B. Guidance Navigation and Control (GNC) API

The GNC API [16] from Intelligent Quads Simulations [21] provides a software collection designed for drone developers with a large amount of documentation and an active community. It implements object detection and avoidance using an interesting principle known as the *Potential Field Method* [23]. This concept was first proposed by O. Khatib in 1985 [24] and suggests creating a potential field where the obstacles generate repulsive potential similar to an energy field. The robot is considered as a particle immersed in this field desiring to reach a destination, while the repulsive force prevents it from getting close to the obstacles [23]. This repulsive potential has an increasing influence when the robot moves closer to the obstacle and decreases upon gaining distance from the obstacle. Let $d_{obst_i}(p)$ be the minimal distance from the robot's position $p := [X, Y, Z, \psi^5]$ to the i 'th obstacle $obst_i$, let k_{obst_i} be a scaling constant and d_0 be the object influence threshold⁶. Then the repulsive potential generated by obstacle i dependent on the robot's position p can be defined as:

$$U_{rep_i}(p) = \begin{cases} \frac{1}{2} k_{obst_i} \left(\frac{1}{d_{obst_i}(p)} - \frac{1}{d_0} \right)^2 & \text{if } d_{obst_i}(p) < d_0 \\ 0 & \text{if } d_{obst_i}(p) \geq d_0 \end{cases} \quad (1)$$

The API uses a two-dimensional 360-degree Hokuyo Lidar (Light Detection and Ranging) [14] sensor for recording the distance data which are being published to a ROS topic. The sensor contains numerous laser rays each of which is able to sense nearby objects and log the data as to where the laser ray is interrupted. The data are then retrieved and parsed by the API in order to perform the aforementioned avoidance algorithm.

⁴A messaging protocol for communicating with(in) drones.

⁵The orientation of the drone in degrees.

⁶The range from which avoidance is triggered.

C. Testing

Several functionalities had to be implemented in order to apply the MRs defined in section III-B2. For instance, functions to measure the travelled distance and passed time, to monitor avoidance behaviour and count the number of avoidance manoeuvres, as well as appropriate configurations to control multiple drones models simultaneously without conflicts. We conducted MT using four carefully defined MRs as well as several alterations of them along with further conventional testing. All test cases have been run multiple times so as to ensure consistency. We generally used an object influence range from 0.35 to 10 metres, a velocity of 1 metre/second, takeoff height of 2 metres and a "waypoint-reached-tolerance" of 0.5 metres (varying parameters have been used for some further testing). Both the distance measurement and avoidance detection mechanisms update once per second, which is accurate enough while only causing minor performance overhead. The launch points and destination targets were positioned 100 metres apart from each other such that the results across various relations are comparable. Test cases included those with no obstacles at all as well as those with obstacles placed in the path connecting the targets. Figure 1 shows a screenshot of the Gazebo world during a test run of R_1 .

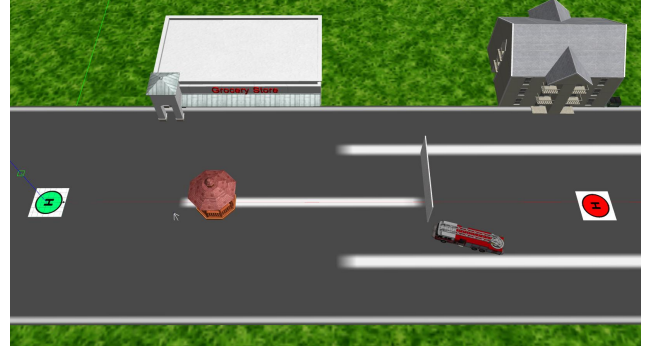


Figure 1: R_1 : One drone moving between two targets, avoiding obstacles.

V. RESULTS AND DISCUSSION

Within this section we present the findings that were revealed during MT using our MRs. We offer a brief interpretation of these results and also demonstrate a few limitations. We also evaluate the suitability of MT as a testing strategy for researchers and software engineers in the AS domain.

A. Consistency without Avoidance

Our first goal was to ensure that the drones themselves show comparable behaviour across equivalent situations without yet paying regard to object avoidance. Fortunately, the relations which did not assess object avoidance (e.g. R_2) showed highly consistent results. Even for varying parameters (e.g. drone velocity, object influence threshold etc.) the resulting distances and durations stayed close together and the relations validated successfully almost all the time. There has been a single

occurrence of a false validation of R_2 , where reviewing the logs showed that one drone took a few seconds to recognise a reached waypoint and move on, causing the durations to differ by approximately 7 seconds. Whenever this delay exceeds a certain amount, the tolerated discrepancy Δ_t will be too strict, leading to a false validation which is precisely the intention of this relation. Since the cause for the delay seemed to be simple overload, we adjusted the previously used parameter $\Delta_t = 3s$ to 10s, with $\Delta_d = 1m$ and have not encountered any failed validations henceforth. The relation R_2 further demonstrated that the object influence threshold is working as intended since the drones were placed approximately 25 metres apart using an object influence threshold of 10 metres, and further tests have been conducted using a threshold of 12 metres. This means that there was only 1 metre of tolerance between the drones along their entire path, though we still have not registered any avoidance manoeuvres. Thus, our testing proved that the drone behaviour and path selection is fairly consistent when leaving object avoidance out of consideration, indicating that testing with object avoidance is reasonable.

B. Challenges of Object Avoidance

During our testing with the MRs that regarded object avoidance, several interesting properties as well as some limitations of the implementation were revealed.

1) *Object Influence Threshold*: The region around the drone where object avoidance is triggered played a decisive role during our testing in regards to object avoidance: The object influence threshold ought to be large enough according to the set drone velocity, such that the drone has enough space to brake before an obstacle. However, this threshold is static and therefore also viable during avoidance, where the velocity is very low. This causes the drone to slowly distance itself from the obstacle until the threshold is reached, move closer again and repeat the process, resulting in a zigzag-shaped path around the obstacle and multiple detected avoidance manoeuvres. For this reason the efficiency of avoidance manoeuvres is severely compromised and the potential field method may not be suitable for robots that are destined to reach a target as quickly as possible. This avoidance issue was revealed by all MRs that assessed object avoidance and e.g. becomes evident immediately from the data logs during the test runs. For instance, the output shown by Figure 2 yields that the number of avoidance manoeuvres is higher than that of obstacles (as seen in Figure 1).

2) *Number of Avoidance Manoeuvres*: The previous figure further illustrates that the drone performed twice as many avoidance manoeuvres on one direction compared to the other one, meaning that different paths have been taken by the drone. The results of R_1 showed that the drone's path $A \rightarrow B$ is generally different from and less optimal than the reverse path $B \rightarrow A$ as in that it requires more avoidance manoeuvres (and therefore more time). The most common avoidance manoeuvre counts were roughly 12 for the path $A \rightarrow B$ and 6 for the path $B \rightarrow A$, which is illustrated in Figure 3 on the basis of 15 test executions using R_1 . The parameters time, distance,

```
...
Waypoint 1 has been reached!
Shortest path to Waypoint:      100.196 m
Distance travelled:             189.520 m
Elapsed wall clock time:       251.815 s
Number of avoidance manoeuvres: 12
...
Waypoint 2 has been reached!
Shortest path to Waypoint:      100.534 m
Distance travelled:             126.710 m
Elapsed wall clock time:       76.002 s
Number of avoidance manoeuvres: 6
...
----- MR Validation -----
MR 1: True
```

Figure 2: Console output snippet of a test run using R_1 .

and avoidance manoeuvres generally showed to correspond to each other, as in that a large number of avoidance manoeuvres indicated more time passed and distance travelled, and vice versa.

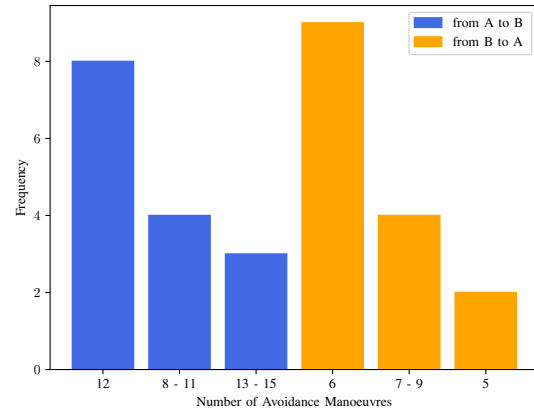


Figure 3: The most frequent avoidance manoeuvre counts using R_1 .

3) *Moving Obstacles*: One of our relations investigated the behaviour of two drones launching from separate starting points, though destined to the same target. Thus, an avoidance manoeuvre by both drones was expected given the drones and their flight parameters are identical. In spite of the drones never crashing and proving consistency of the Lidar sensor, there have been occasions of rather unfortunate avoidance behaviour. The drones would keep symmetrically avoiding each other towards the same direction and continue to move further and further away from the target. Consequently, one test run took over 13 minutes for both drones to complete their paths, while it took between one and three minutes in most other test executions. The total time taken by each drone to perform its mission is shown by Figure 4 based on 15 example test runs. In spite of the highly inconsistent observations as well as the overall fluctuating durations, it was detected that the durations required by each drone stayed highly similar within a single test run. The occasional deviations mostly traced back to one drone taking a tiny amount of time longer to take off or to detect its waypoint, or ending up ahead/behind the other drone due to an asymmetrical avoidance manoeuvre,

which then has a major influence on the subsequent behaviour. Overall, it was surprising to see how random the drones behave albeit being controlled exactly the same way.

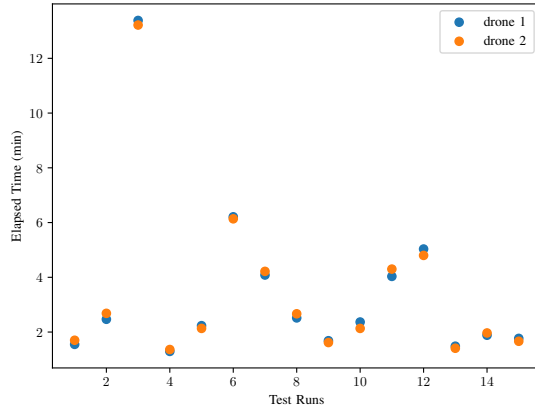


Figure 4: The durations taken by each drone to perform a mission with intersecting paths.

4) *Distance Vector Conflicts*: We have observed a bug where a drone would avoid an obstacle and then suddenly start registering to have reached waypoints and ultimately land, despite certainly not being close to any waypoint. The function that checks whether a waypoint has been reached seemed to work just fine upon investigation and adjusting the control loop rate seemed to partly alleviate the issue, surprisingly. We have discussed this bug with one of the API's developers, Sahasrajit A., who recommended suspending waypoint pursuit whenever object avoidance is active. This is because the drone might be confused as to where to go during avoidance due to conflicting distance vectors (the avoidance algorithm demands to move away from the obstacle whereas the waypoint pursuit instruction concurrently orders the drone to move forward). With the appropriate implementation changes in place, we have been unable to reproduce this issue henceforth.

5) *Limitations of the Sensor*: The relation R_1 further demonstrated that certain, complex shapes of obstacles may pose a challenge for the two-dimensional sensor. Consider a fire truck model (visible in Figure 1) which has a folded aerial ladder mounted on its roof, increasing the vehicle's height at its centre. When the drone is travelling at the height of the truck's ladder, it is possible that the drone moves too close and slightly touches the truck with its standoffs. Such an incident occurred twice and essentially traces back to the two-dimensional nature of the sensor as well as the obvious fact that the drone model is larger in height than the line length of the laser rays⁷. Increasing the object influence threshold will alleviate this issue, while altering the sensor would probably be a better solution. Further testing has again brought up the two-dimensionality of the sensor in the case of higher wall heights. The most efficient way of avoiding this wall would clearly be to increase the flight height where the wall is no longer sensed and continue to move forward, avoiding the wall

⁷Laser Line Length [accessed Apr-2022]

from above. However, the sensor is unaware of the wall having only insignificant height and is going to demand the drone to move left or right until the wall ends. Metamorphically, this difficulty can be revealed through an extensive amount of time taken and/or distance travelled in spite of a small distance of the shortest path. The purely horizontal behaviour is not only inefficient in this case, but may also lead to a complete failure in a situation where the drone is completely surrounded by obstacles. It becomes more and more evident that a single, two-dimensional sensor is probably not sufficient for a real-world application and vertical avoidance capabilities would be desirable. It was also found that the number of avoidance manoeuvres may be useful for evaluating path efficiency: The less avoidance manoeuvres are performed, the more linear the path is going to be which equals a higher efficiency. This dependency could be utilised for a Machine Learning approach where the drone learns to choose more efficient paths. An AI drone could sense that the path $B \rightarrow A$ required notably fewer manoeuvres while connecting the same points A and B, and hence choose this path over the path $A \rightarrow B$ in the future. Besides, the fact that the most frequent avoidance manoeuvre counts were far apart from each other across the paths $A \rightarrow B$ and $B \rightarrow A$ shows that obstacle positions along with the drone's launch point heavily influence the path chosen by the drone. The results of using other drones as moving obstacles underlined that the autonomous system in combination with the used algorithm is not stable as in that it does not behave essentially the same way for exactly the same situation. In fact, it would require intelligent path selection or waiting behaviour in order to attain good efficiency in such missions. This could be achieved by an approach where drones communicate with each other and adjust their paths according to the other drone's destination and position. For multiple reasons, vertical avoidance and compatibility with all object shapes would be highly desirable, hence we would recommend inducing a three-dimensional sensor. With such a sensor in place, one could conduct a wide range of further tests using a vertical obstacle course.

C. Expedience of Metamorphic Testing

Applying MT to this domain has significantly improved our domain knowledge and brought up several important aspects about the system under test. Even during the process of MR identification, the potential of this testing technique in AS has already started to show: during the construction of R_1 , it was initially planned to demand that,

$$f(x_1)_{ac} = f(x_2)_{ac} \quad (2)$$

However, the nature of the avoidance algorithm suggests that the drone may not choose an identical path for both path directions, resulting in potentially varying avoidance manoeuvre counts despite fundamentally correct behaviour. This aligns with the reasoning for a longer travel distance expected by the unmanned aerial vehicle (UAV) example for the transformation IV rule presented by Li et al. [25]. Thus, formulating a weak MR actually proved to be useful upon

realising its deficiencies. This demonstrates how extensively MRs ought to be thought through and how much there is to be gained from concerning oneself with MT in this domain.

VI. RELATED WORK

Ever since the proposition of MT, researchers and engineers have been applying the technique to real-life applications across different fields such as compilers [26], search engines [27], computer graphics [28], machine learning [29], autonomous systems [5], [30] and numerical analysis [31]. A major survey in MT was conducted by Segura et al. in 2016 [1] and updated in 2020 [2]. Segura et al. studied the trends in MT and their results indicate that MT is a thriving topic with an increasing amount of contributions and concludes that it is an effective testing technique for alleviating the oracle problem. When one analyses Figure 5 and compares it to the older survey of 2016, one observes that applications in simulation and modelling have decreased by 4%, whereas applications in bioinformatics have increased by 3%. There has also been a 2% increase in the autonomous vehicles domain even though it was not previously obtainable in the former survey [1] at all. This leads to the assumption that the number of applications in this domain used to be negligibly small until 2016, but has been rising in the subsequent years. This has partly motivated out work since there has been a very limited amount of studies on MT in autonomous systems and their simulation hitherto.

Segura et al. [2] claim that it is essential to comprehend that applications among different domains may ask for structural adaptation within the concept, yielding different approaches to MT. For instance, researchers in compiler testing proposed that removing dead code should not influence the functionality of the compiler-generated code. Their idea is to simply run an initial test case, remove some dead code and re-run the program (“next” test case) treating any observed changes as a failure. This approach has been applied in validating the GCC (GNU Compiler Collection) and LLVM (Low-Level Virtual Machine) [26]. Another survey followed in 2018 by Chen et al., which highlights some vital challenges and unveils the most promising opportunities for future research on MT [9]. The authors recommended for researchers who develop software engineering methods that may require a test oracle to consider MT during development, since this may alleviate the oracle requirement. Liu et al. [13] provided answers to the question of how effectively MT alleviates the oracle problem, conducting an empirical analysis where MT was performed by different testers and on multiple programs. They conclude that MT does generally alleviate the oracle problem, and argue that the diversity of MRs is much more important than their quantity.

In autonomous system simulations, Lindvall et al. [5] have presented a model-based testing approach combined with MT principles for testing of autonomous drones. In fact, their work posed a major motivation for our project. The authors presented a number of equivalences for their AI drone which represent an outstanding base for finding MRs and constructing test cases. The simulation environment for their project

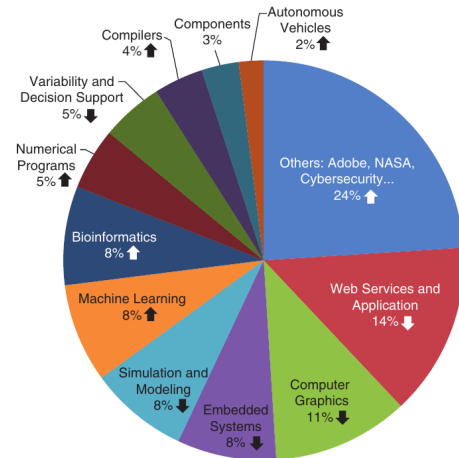


Figure 5: MT application domains (2020) [2]

was implemented using Unity 5. Their approach showed potential for identifying corner cases and thereby increased their confidence in the autonomous system. They proposed to apply the testing to more MRs, use other drones as moving obstacles as well as to add physical conditions such as wind.

In conclusion, the impressive amount of recent contributions to MT suggests that the testing strategy is being recognised as a promising approach for alleviating the oracle problem, although there are only few papers to be found concerning MT in autonomous system simulations. Still, the few applications in autonomous systems showed impressive results which indicate that MT may indeed be an adequate technique for this domain [5], [30].

VII. CONCLUSION

We have provided an overview of MT as well as an implementation of this testing technique in autonomous drone simulations. In adherence with specific properties of effective MRs, multiple relations have been carefully defined and utilised for testing. We have demonstrated what Lindvall et al. [5] proposed in their future work section, i.e. to apply multiple MRs and perform MT in a simulation with multiple drones. We posit that despite a seemingly simple algorithm, complex and diverse observations of drone behaviour may occur. We have also demonstrated a dependency between drone velocity and the object influence threshold, which is a crucial property when utilising the potential field method and causes an unfortunate avoidance behaviour in many cases.

To conclude, through conducting MT in our simulation we have been able to deduce several essential properties of the system under test that we were previously unfamiliar with. With this, we have investigated some issues that only a few previous studies have dealt with. Each of the implemented metamorphic relations has revealed vital details about the flaws and behaviours of the used software and object avoidance strategy. Despite the algorithm technically working as intended (for particles), there may be other choices that are more useful to our setting. Due to the complexity of this domain and the

typically limited quality of documentation, we trust that this paper and the technical insight it provides may be helpful to other developers. Concerning ourselves with MT in this domain has certainly improved our own understanding of the testing strategy as well as the domain. Both the testing process and the results indicated great applicability for our setting which is why we wish to recommend MT to developers in autonomous systems. In particular, we believe that invoking MT as a complementary addition to conventional software testing is going to have a very positive influence in nearly all cases.

In the future, for the potential field method, we plan to invoke dynamic object influence thresholds according to the drone velocity, such that the threshold increases when the drone accelerates and decreases when it brakes. As a consequence, the efficiency of an avoidance manoeuvre could be massively improved and the observed zigzag-shaped avoidance path could be eliminated. We would also like to incorporate intelligent path planning (selection) by trying an algorithm of a different nature like the Bug algorithms [23], where avoidance paths are generated around obstacles. AI drones could learn to choose the most efficient path according to the amount of required avoidance manoeuvres, whereas for other drones as obstacles they could simply communicate with each other and adjust their paths according to the other drone's destination and position, similar to what Chen et al. [32] have done. One could then utilise shape matching algorithms to evaluate the quality of chosen paths and use this for MRs, as Lindvall et al. [5] have done similarly.

In the same degree, we plan to implement a 3D obstacle detection method that enables the detection of obstacles all around a test drone for instance, thereby ensuring that obstacle avoidance manoeuvring can happen in a more realistic manner as would be obtainable when the AS is in operation. With a reliable avoidance behaviour and our functionality of detecting a manoeuvre, testers could start to become stricter with the expected number of avoidance manoeuvres in MRs and thus really narrow down the inconsistencies of the used avoidance algorithm.

REFERENCES

- [1] S. Segura, G. Fraser, A. B. Sanchez, and A. Ruiz-Cortés. A survey on metamorphic testing. *IEEE Transactions on Software Engineering*, 42(9):805–824, 2016.
- [2] S. Segura, D. Towey, Z. Q. Zhou, and T. Y. Chen. Metamorphic testing: Testing the untestable. *IEEE Software*, 37(3):46–53, 2020.
- [3] T. Y. Chen, P. Poon, and X. Xie. Metric: Metamorphic relation identification based on the category-choice framework. *Journal of Systems and Software*, 116:177 – 190, 2016.
- [4] Elaine J. Weyuker. On Testing Non-Testable Programs. *The Computer Journal*, 25(4):465–470, 11 1982.
- [5] M. Lindvall, A. Porter, G. Magnusson, and C. Schulze. Metamorphic model-based testing of autonomous systems. In *2017 IEEE/ACM 2nd International Workshop on Metamorphic Testing (MET)*, pages 35–41, 2017.
- [6] Hans-Peter Schöner. Simulation in development and testing of autonomous vehicles. In Michael Bargende, Hans-Christian Reuss, and Jochen Wiedemann, editors, *18. Internationales Stuttgarter Symposium*, pages 1083–1095, Wiesbaden, 2018. Springer Fachmedien Wiesbaden.
- [7] Tom P. Huck, Christoph Ledermann, and Torsten Kröger. Testing robot system safety by creating hazardous human worker behavior in simulation. *IEEE Robotics and Automation Letters*, 7(2):770–777, 2022.
- [8] T. Y. Chen, S. C. Cheung, and S. M. Yiu. Metamorphic testing: A new approach for generating next test cases, 2020.
- [9] T. Y. Chen, F. Kuo, H. Liu, P. Poon, D. Towey, T. H. Tse, and Z. Q. Zhou. Metamorphic testing: A review of challenges and opportunities. *ACM Comput. Surv.*, 51(1), 2018.
- [10] J. Mayer and R. Guderlei. An empirical study on the selection of good metamorphic relations. volume 1, pages 475–484, 01 2006.
- [11] T. Chen, D. Huang, T.H. Tse, and Z.Q. Zhou. Case studies on the selection of useful relations in metamorphic testing. 11 2004.
- [12] Z. Hui and S. Huang. A formal model for metamorphic relation decomposition. In *2013 Fourth World Congress on Software Engineering*, pages 64–68, 2013.
- [13] H. Liu, F. Kuo, D. Towey, and T. Y. Chen. How effectively does metamorphic testing alleviate the oracle problem? *IEEE Transactions on Software Engineering*, 40(1):4–22, 2014.
- [14] Niroj Pokhrel. Drone Obstacle Avoidance and Navigation Using Artificial Intelligence. Master's thesis, Aalto University. School of Science, 2018.
- [15] T. Ryan, H. Willee, K. Hester, R. Roche, P. Barker, and P. Nicolao. DroneKit Python API. <https://github.com/dronekit/dronekit-python>, 2014. [Online: Github repository; Jan-2022].
- [16] E. Johnson, S. Ananth, F. Furrer, M. Burri, M. Kamel, J. Nikolic, and M. Achtelik. Intelligent Quads Drone Development Software Collection. https://github.com/Intelligent-Quads/iq_gnc, 2019. [Online: Github repository; accessed Jan-2022].
- [17] S. P. de Melo, M. da Silva Neto, G. da Silva Neto, José Lima da Silva, Jorge Lima da Silva, P. Teixeira, N. Teixeira, J. M. Xavier, and V. Teichrieb. Analysis and comparison of robotics 3d simulators. In *2019 21st Symposium on Virtual and Augmented Reality (SVR)*, pages 242–251, 2019.
- [18] Anna Konrad. Simulation of mobile robots with unity and ros : A case-study and a comparison with gazebo. 2019.
- [19] A. Staranowicz and G.L. Mariottini. A survey and comparison of commercial and open-source robotic simulator software. In *Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '11*, New York, NY, USA, 2011. Association for Computing Machinery.
- [20] A. Tridgell, R. Mackay, and P. Barker. ArduPilot Project. <https://github.com/ArduPilot/ardupilot>, 2010. [Online: Github repository; accessed Jan-2022].
- [21] E. Johnson, F. Furrer, M. Burri, M. Kamel, J. Nikolic, and M. Achtelik. Intelligent Quads Simulations. https://github.com/Intelligent-Quads/iq_sim, 2018. [Online: Github repository; accessed Jan-2022].
- [22] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Ng. ROS: an open-source Robot Operating System. volume 3, 01 2009.
- [23] M. I. Ribeiro. Obstacle avoidance. 2005.
- [24] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. In *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, volume 2, pages 500–505, 1985.
- [25] Rui Li, Huai Liu, Guannan Lou, Xi Zheng, Xiao Liu, and Tsong Yueh Chen. Metamorphic testing on multi-module uav systems. In *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pages 1171–1173, 2021.
- [26] V. Le, M. Afshari, and Z. Su. Compiler validation via equivalence modulo inputs. *SIGPLAN Not.*, 49(6):216–226, 2014.
- [27] Z. Q. Zhou, S. Xiang, and T. Y. Chen. Metamorphic testing for software quality assessment: A study of search engines. *IEEE Transactions on Software Engineering*, 42(3):264–284, 2016.
- [28] A. F. Donaldson, H. Evrard, A. Lascu, and P. Thomson. Automated testing of graphics shader compilers. *Proc. ACM Program. Lang.*, 1(OOPSLA), 2017.
- [29] X. Xie, J. W. K. Ho, C. Murphy, Gail K., B. Xu, and T. Y. Chen. Testing and validating machine learning classifiers by metamorphic testing. *Journal of Systems and Software*, 84(4):544 – 558, 2011.
- [30] S. Wang and Z. Su. Metamorphic testing for object detection systems, 2019.
- [31] Z. Q. Zhou, D. H. Huang, T.H. Tse, Z. Yang, H. Huang, and T. Y. Chen. Metamorphic testing and its applications. *ISFST 2004*, 2004.
- [32] Jiajia Chen, Rui Zhang, Wei Han, Wuhua Jiang, Jinfang Hu, Xiaoshan Lu, Xingtao Liu, and Pan Zhao. Path planning for autonomous vehicle based on a Two-Layered planning model in complex environment. *Journal of Advanced Transportation*, 2020:6649867, November 2020.