Noname manuscript No.
(will be inserted by the editor)

Testing the Robustness of Controllers for Self-Adaptive Systems

Javier Cámara · Rogério de Lemos · Nuno Laranjeiro · Rafael Ventura · Marco Vieira

Received: date / Accepted: date

Abstract Self-Adaptive systems are software-intensive systems endowed with the ability to respond to a variety of changes that may occur in their environment, goals, or the system itself, by adapting their structure and behavior at runtime in an autonomous way. Controllers are complex components incorporated in self-adaptive systems, which are crucial to their function since they are in charge of adapting the target system by executing actions through effectors, based on information monitored by probes. However, although these controllers are becoming critical in many application domains, so far very little has been done to assess their robustness. In this paper, we propose an approach for evaluating the robustness of controllers for self-adaptive software systems, aiming to identify faults in the design of these controllers. Our proposal considers the stateful nature of the controller, and identifies a set of robustness tests, which includes the provision of mutated inputs to the interfaces between the controller and the target system (*i.e.*, probes). The feasibility of the approach is evaluated on Rainbow, a framework for architecture-based self-adaptation, and in the context of the Znn.com case study.

Keywords
Robustness Testing, Controller, Self-Adaptive System, Autonomic System

J. Cámara
Carnegie Mellon University, USA
Tel.: +1-412-268-1601
Fax: +1-412-268-3455
E-mail: jcmoreno@cs.cmu.edu

R. de Lemos
University of Kent, UK
E-mail: r.delemos@kent.ac.uk

N. Laranjeiro, R. Ventura, M. Vieira
University of Coimbra, Portugal
E-mail: {cnl,ventura,vieira}@dei.uc.pt

1 Introduction

One of the main traits of a self-adaptive software system, compared to any other kind of system, is its ability to deliver its services in spite of changes that may occur in the system, its environment or even in its goals. A key component that enables self-adaptive systems to handle changes at run-time (*e.g.*, repairing anomalies and improving operation) is a controller that relies on a feedback control loop for managing adaptations [3] by executing actions through system-level effectors on the target system, based on information monitored by probes. In the context of complex software systems, these controllers typically implement the traditional *sense-plan-act* architectures. An example of such controllers is the MAPE-K model, which includes four distinct operational stages, namely, monitoring, analysis, planning and execution [14]. Despite major achievements in the area, existing approaches in autonomic systems and self-adaptation do not systematically address the need to determine if a self-adaptive system can deliver a service that can justifiably be trusted when facing changes (*i.e.*, that it will be *resilient* [17]). This lack of assurances is an issue that has hampered the widespread adoption of self-adaptive systems, which are often regarded as unreliable by industry. A major problem associated with the provision of evidence is the combinatorial nature of the stateful aspects of a controller and the changes that may affect the system being controlled. Since the different operational stages in the feedback control loop should be functionally independent from each other, a change might have a different impact on the controller depending on the state of the controller. Moreover, if the controller is expected to act upon a change when it occurs,

there is a wide range of issues that needs to be considered when producing the appropriate action, including the place in which the change has occurred, the type and the frequency of the change, and whether it can be anticipated [2]. These factors have to be considered regarding the provision of assurances about the services to be delivered by the target system. Hence, novel techniques need to be devised in order to uncover potential faults in the controller.

The present paper describes an approach for evaluating the robustness of controllers for self-adaptive systems by abstracting away, in a first instance, from the state of the target system being controlled. The rationale behind this is the fact that the complexity associated with these controllers is such that we need first to devise novel means for evaluating the core logic that enables adaptation, before exploring the ensemble target system plus controller. Moreover, if the robustness evaluation is performed on the ensemble, some of the controller faults could be masked by the target system, or their effects upon the system could be more difficult to analyze. Hence the decision to define an approach that can be used in the robustness evaluation of different controllers, assuming that the core logic of the different operational stages is basically the same on the different controllers [10]. In such a way, we restrict the robustness tests in our approach to the inputs of the controller, which are characterized by the probes. Although the proposed approach abstracts away from the target system, we need to consider the stateful aspects of the controller, which are related to its different operational stages.

The primary contribution of this paper is the definition of an approach for evaluating the robustness of controllers, which part of a bigger initiative that is looking into the resilience evaluation of self-adaptive systems. Our proposal considers the stateful nature of the controller by defining how the controller interface should be tested according to a target system changeload [1,6], and the operational stage of the controller. To achieve our goal, the approach defines a set of mutation rules that should be applied to the inputs of the controller, a tailored version of a classification of the different controller failure modes, and an experimental setup and testing procedure that is specific to self-adaptive systems. A preliminary evaluation [5] of the feasibility of our approach was carried out using the Rainbow framework, which consists of a controller that supports architecture-based self-adaptation [10], and in the context of a simplified version of the Znn.com case study [8]. Experimentation using Rainbow is very convenient, since its software has been widely available, its structure facilitates access to its internal components, its design is amenable to the injection of faults, and the logs Rainbow produces are suitable for analyzing the effects of the injected faults upon the controller. The present paper extends our preliminary study by reporting on an exhaustive evaluation of the approach on a full-fledged deployment of Rainbow/Znn.com, including extensive tests carried out on a comprehensive set of probes implemented using different technologies.

The rest of this paper is structured as follows. Section 2 provides some background on self-adaptive systems and related work in the area of robustness testing. Section 3 introduces the Znn.com case study, which is used throughout the paper for illustrating the proposed approach. Section 4 describes our approach that is focused specifically on evaluating the robustness of controllers for self-adaptive systems. Section 5 presents the experimental results obtained from the evaluation of our approach. Finally, Section 6 concludes the paper and indicates future research directions.

2 Background and Related Work

The run-time management of increasingly complex software-intensive systems has become a central concern in Software Engineering over the last few years [7,19]. Specifically, a major issue in the area is concerns achieving conformance to functional and non-functional requirements in a dependable and cost-effective manner despite the influence of changes that may affect the system, its environment, and system goals.

One of the seminal works addressing this concern was IBM's Autonomic Computing initiative [14], which introduced a layer implementing what is known as the MAPE-K control loop to Monitor, Analyze, Plan, and Execute adaptation (with a Knowledge base that supports the different activities in the control loop) for the purpose of managing a target system. In particular, some successful approaches that rely on this closed-loop control paradigm for self-adaptation exploit architectural models for high-level reasoning about the target system under management [10,23]. In particular, Rainbow [10] is a framework which provides a base of reusable infrastructure that can be applied to a wide range of systems through customization. The framework defined by Rainbow includes mechanisms for monitoring a target system and its environment (using the observations for updating the architectural model of the target system), detecting opportunities for improving the system's quality of services (QoS), deciding the best course of adaptation based on the state of the system. Section 4.4 provides further details about the Rainbow framework, which is used for the experimental validation of our approach.

2.1 Resilience Evaluation in Self-Adaptive Systems

Despite the fact that research in the field of autonomic and self-adaptive systems is relatively new, there are already some contributions regarding their provision of assurances. However, the applicability of these contributions has been focused on the ensemble target system plus controller. To the

best of our knowledge, no approaches have been proposed regarding the evaluation of controllers, although there is already some ground work pointing in this direction [7, 19].

One of the areas that are related to that of resilience evaluation is that of resilience benchmarking, which encompasses techniques from previous efforts in performance benchmarking [11], dependability benchmarking [13], and security benchmarking [20], due to its inherent relation to performance, dependability and security. Comparing to established benchmarks, a resilience benchmark may be specified following the same basic approach, but comprising a wide-ranging changeload (which will include, but will be not limited to, faults), as well as resilience metrics [1].

Other approaches deal with resilience evaluation through quantitative analysis using probabilistic model-checking [4], considering the system environment as the only source of change and leaving out changes that are internal to the system. The cited approaches quantitatively measure resilience in the self-adaptive system when facing changes either internal or external to the system. However, they do not deal with an additional source of problems from the perspective of resilience, which are robustness issues addressed by the techniques presented in the current paper.

2.2 Robustness Testing

Robustness testing consists in stimulating a system with erroneous input conditions with the goal of triggering internal errors. This allows testers to differentiate systems according to the number and type of errors uncovered and provides developers with information to solve or wrap the identified problems [22].

Ballista [15] uses a set of tests that combine acceptable and exceptional values on calls to kernel functions of operating systems. The parameter values used in each invocation are randomly extracted from a set of predefined tests and for each parameter a set of values of a certain data type is associated. Each operating system is classified in terms of its robustness and according to a predefined scale (the CRASH scale [15]) that distinguishes several failure modes.

Initially, Ballista was developed for POSIX APIs (including real time extensions). Further work has been developed to adapt it to Windows operating systems [28]. In that study the authors present the results of executing Ballista generated exception handling tests over several functions and system calls in Windows 95, 98, CE, NT, 2000, and Linux. The authors were able to trigger system crashes in Windows 95, 98, and CE. The other systems also revealed robustness problems, but not complete system crashes.

MAFALDA (Microkernel Assessment by Fault injection AnaLysis and Design Aid) [25] is a tool that enables the characterisation of the behaviour of microkernels in the presence of faults. Fault injection is performed at two levels: in the parameters of system calls, and in the memory segments holding the target microkernel. However, only the former is relevant when the goal is robustness testing.

The robustness testing techniques have been applied not only at the operating system level but also at the middleware layer and targeting different types of systems. The problem of robustness testing of high availability middleware is discussed in [21]. The paper presents a testing framework that integrates previous testing techniques (*e.g.*, scenario-based testing and test result classification). The case study conducted on OpenAIS (an open implementation of the Application Interface Specification (AIS) provided by the Service Availability Forum) showed that simple techniques can identify robustness problems. However the implementation of more complex techniques is required since these are able to find faults not detected by the simple ones.

Ballista was also adapted to be applied to middleware systems. In particular, [24] studies the robustness of various CORBA ORB implementations. In this case, the failure modes were adapted to better characterize the CORBA context and the authors were able to reveal several issues in the middleware being tested.

In [18] we propose an experimental approach for the robustness evaluation of JMS middleware. The technique is applied successfully to three major JMS middleware providers exposing serious robustness problems, including severe security issues, which also highlights the importance of the application of robustness testing to real-world systems.

The abovementioned works implement robustness testing approaches that do not consider the state of the system under test. In [9] the impact of state on robustness testing of a safety-critical operating system (OS) is investigated by including the OS state in test cases definition. Although system-specific, results show that the state can play an important role in testing since they are able to cover more cases when compared to the traditional approaches.

An approach for robustness testing method of stateful Web services, modelled with Symbolic Transition Systems, is presented in [27]. A test case generation method is proposed using unusual values and replacement and additions of operation names. States are transversed using different operations and starting from a system specification which, depending on the system being tested, may not always be available. The authors assume that messages sent and received are only SOAP messages and suggest that a Web service could be considered as a grey box from which any type of message could be observed, increasing the potential of the technique.

In [5] we present an approach to evaluate the robustness of controllers for self-adaptive software systems, aiming at the identification of design faults. The approach is based on a set of robustness tests that include the provision of mutated inputs to the interfaces between the controller and the target

system (*i.e.*, probes). The feasibility of the approach is evaluated in the context of Znn.com, a case study implemented using the Rainbow framework for architecture-based self-adaptation.

3 Case Study

To illustrate our approach for robustness testing, we use the Znn.com case study [8], which is implemented using Rainbow, and is able to reproduce the typical infrastructure for a news website. It has a three-tier architecture consisting of a set of servers that provide contents from backend databases to clients via front-end presentation logic. Architecturally, it is a web-based client-server system that satisfies an N-tier style, as illustrated in Figure 1. The system uses a load balancer to balance requests across a pool of replicated servers, the size of which can be adjusted according to service demand. A set of client processes makes stateless requests, and the servers deliver the requested contents (*i.e.,* text, images and videos).

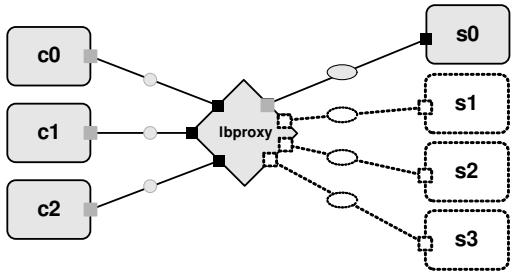

Fig. 1 Znn.com system architecture

The main objective for Znn.com is to provide content to customers within a reasonable response time, while keeping the cost of the server pool within a certain operating budget. It is considered that from time to time, due to highly popular events, Znn.com experiences spikes in requests that it cannot serve adequately, even at maximum pool size. To prevent losing customers, the system can provide minimal textual contents during such peak times, instead of not providing service to some of its customers. Concretely, there are two main quality objectives for the self-adaptation of the system: *(i)* performance, which depends on request response time, server load, and network bandwidth, and *(ii)* cost, associated to the number of active servers.

In the case of Znn.com, Rainbow is capable of analysing trade-offs among the different objectives, and execute different adaptations according to the particular run-time conditions of the system. For instance, when response time becomes too high, the system should increment server pool size if it is within budget to improve its performance; otherwise, servers should be switched to textual mode (start serving minimal text content) if cost is near budget limit.

4 Approach

Our approach for robustness evaluation of controllers in a self-adaptive software system considers the model depicted in Figure 2. The *environment* consists of all non-controllable elements that determine the operating conditions of the system (*e.g.*, hardware, network, physical context, etc.). Regarding the system itself, we distinguish two main subsystems: a *target system*, which interacts with the environment by monitoring relevant variables associated with operating conditions, and a *controller* that manages the target system, driving adaptation whenever it is required. Concretely, the controller carries out its function by: *(i)* monitoring the target system and its environment by means of *probes* that provide information about the value of relevant variables, *(ii)* deciding if the current state of the target system and environment demands adaptation, and if this is the case, *(iii)* applying a sequence of control actions through system-level *effectors*.

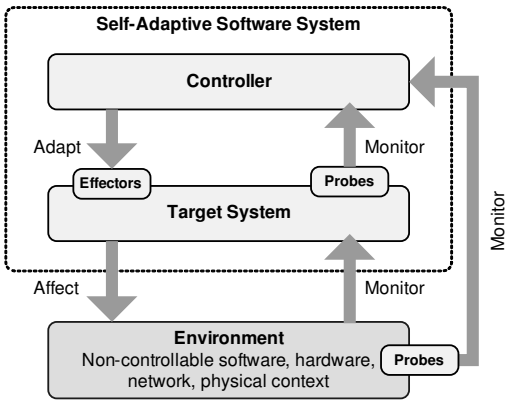

Fig. 2 Self-adaptive software system

In this work, we focus exclusively on the robustness of the controller, *i.e.,* we modify the probes' inputs into the controller with the intent of evaluating how robust is the controller regarding changes that may affect its interface when exceptional input is provided. The controller is considered as a stateful entity, regarding evaluation purposes, since for the same input, the controller's internal state may influence its output. In order to tackle this issue, we consider input mutation during the different operation stages of the controller [1] to create an appropriate context for evaluating its robustness. The key elements of our approach are: *changeload*, which is a set of representative change scenarios, where changes are

[1] Specifically, during analysis, planning, and execution. Monitoring is transversal to the rest of the activities in the MAPE-K loop.

based on controller input mutations; *failure mode classification*, that characterizes the run-time behaviour of the controller while the target system is running in the presence of the changeload; and *robustness tests*, the mutation rules that are applied to the input probes into the controller.

In the following, in addition of describing the key elements of our approach, we also present how robustness tests are performed at the controller interface by mutating the inputs provided by the probes. To exemplify the principles of our approach, we have instantiated it into Rainbow [10].

4.1 Changeload Model

This section describes the proposed model for the changeload, presenting the definitions adopted for the fundamental concepts that form the basis of its structure.

Definition 1 (Change Type) A change type is a tuple (src, m, A) that characterises a change, where:

- src identifies the source probe type where a mutation rule is applied,
- m identifies the mutation rule applied on probe input,
- $A = \langle a_1, \ldots, a_n \rangle$ (possibly empty) is a vector of attributes that holds specific information about the mutation rule.

Example 1 In Znn.com, consider the change "Set an invalid timestamp date on a response time probe (type ClientProxyProbeT)". A possible change type definition for this would be:

invalidDateCPP_CT= (ClientProxyProbeT, TSInvalidDate, ⟨date⟩)

Definition 2 (Change) Given a set of change types CT, a change is a tuple $(ct, srcinst, V_A, ti, d)$ that corresponds to an instantiation of a change type, where:

- $ct = (src, m, A) \in CT$ determines the change type to be instanced as a change,
- $srcinst$ is the probe instance that is the source of change (*i.e.*, in which the input is mutated),
- $V_A = \langle v_{A1}, \ldots, v_{An} \rangle$ is a vector of attribute values instantiating the attributes in A,
- $ti \in \mathbb{R}_0^+$ determines the time instant in which the change is triggered,
- $d \in \mathbb{R}^+$ is the duration associated with the change.

It is worth observing that while some specific changes may be transient, impacting the controller's input during a particular amount of time, in the definition above duration can be considered equal to ∞ if the change is permanent.

Example 2 If we consider the change type described in Example 1, a possible instantiation of it could be:

(invalidDateCPP_CT, ClientProxyProbe1, ⟨'2/29/1985'⟩, 10, 2)

The systematic identification and classification of change types is fundamental to support the definition of change scenarios, which is discussed in the next paragraphs.

The main base concept in our changeload model is the *scenario*. A scenario is a postulated sequence of events that captures the state of the system and its environment, system goals [2], and changes affecting all the aforementioned elements. It is defined in terms of state (system and environment) and changes applied to that state.

Definition 3 (Scenario) A scenario is a tuple (wl, oc, C), where:

- wl represents the workload, that is, the amount and type of work assigned to the system (not necessarily static),
- oc are the operational conditions of the system (including software and hardware resources needed for the system to perform its service),
- C is a set of changes applied to controller input in the presence of the workload and operational conditions.

Based on the definition above, a *change scenario* is one which includes a non-empty set of changes ($C \neq \emptyset$).

Definition 4 (Changeload) A changeload is a set of change scenarios.

4.2 Controller Failure Modes

The robustness of a controller for a self-adaptive system can be classified according to an adapted version of the CRASH scale [16], which distinguishes the following failure modes:

1. **Catastrophic**: the whole controller crashes or becomes corrupted (this might include the OS or machine on which the controller is running). No output is produced.
2. **Restart**: the controllers execution hangs and may not issue any output commands, or send always the same command, within the worst case execution time associated with the adaptation cycle. The controller needs to be externally re-booted.
3. **Abort**: abnormal behavior in the controller occurs due to an exception raised at run-time inside of the controller.
4. **Silent**: the controller fails to acknowledge an error, for instance by signalling an exception, which causes the controller to continue operating improperly.
5. **Hindering:** the controller fails to return a correct error code, which may hinder error recovery. The difference between a silent failure and this case is that, here an error is acknowledged by the controller but the returned error code is incorrect.

[2] For the sake of simplicity, in this paper we abstract away from system goals, which are not required to deal with robustness evaluation of the controller.

In particular, it is worth observing that the tailored version of the CRASH scale for controllers in self-adaptive software systems includes an specific adaptation which is related with time (2).

4.3 Robustness Tests

The basis of the proposed approach for evaluating the robustness of controllers for self-adaptive software systems relies on stimulating the interface of the controller, which consists of probes that monitor both the target system and its environment (see Figure 2). For evaluating how robust is the controller, regarding changes that my affect its interface, the probes' inputs into the controller are modified according to a comprehensive set of mutation rules. Moreover, since the inputs of these probes may affect the different stages of a MAPE-K control loop, the evaluation needs to consider the controller as stateful. Although for evaluating the robustness of a controller we are able to abstract away from the application (target system), we nevertheless use the application to drive the evaluation.

4.3.1 Mutation Rules

The set of robustness tests performed is automatically generated by applying a set of predefined mutation rules to the messages sent by probes, which characterizes the monitoring stage of the controller. Although concrete message formats and additional elements may exist depending on the case, the basic input supplied by probes to the controller typically consists of three basic elements: *(i)* an identifier of the variable being monitored, *(ii)* the actual value for the variable, and *(iii)* a timestamp that provides a temporal context for the variable being monitored. For example, in the case of Rainbow, the kind of input received by the controller consists of simple messages encoded as text strings with the following format:

```
[ timestamp ] variable_name : variable_value
```

Based on this general description of probe input, we propose a set of rules (Table 1), which have been defined based on previous works on robustness testing [16, 29, 26], and explore limit conditions that are typically the source of robustness problems.

4.3.2 Probe Usage Categories

The effect of applying mutation rules on the outputs generated by the probes may manifest in different ways (or not manifest at all) in the controller, depending on its internal state. This results from the stateful nature of the controller, which may use different inputs and in a different way, depending on its operation stage (*i.e.*, analysis, planning, or execution). Changes in the internal operation stage of the controller are also induced by input obtained from probes.

Table 2 distinguishes different probe categories, according to their use in the different operation stages of the controller. Different robustness issues may arise in the controller, depending on the particular stage/probe in which mutation rules are applied, even if the set of mutations rules applied are the same. The same probe can belong to different usage categories and be used during different stages in the controller. We consider the controller to be a gray box, while its different operation stages are black boxes on which probe mutation is applied. For the time being, we assume that each of these black boxes are stateless, even if that is not the case as far as the target system is concerned. The different stages in the controller are sequential, while monitoring is transversal to all of them.

4.4 Testing Procedure

As discussed previously, inputs to the Rainbow controller are delivered with the use of probes, which provide important system and environment information such as experienced response time, network latency, or server load. Robustness testing focuses on the controller's input points (*i.e.,* the probe information). Therefore, a complete robustness experiment must include a set of tests that focuses precisely on the information provided by each of the input probes.

Figure 3 represents the complete experimental procedure and, as we can see in the figure, each experiment includes several tests, each one focusing on a given probe. For each probe (which, at runtime is continuously delivering information to the controller under test) we apply a single change for each probe data sample. However, we apply (in the subsequent probe data samples) the same change for a given period of time, which potentially gives us the possibility of further disturbing the system under test.

Each robustness test focuses on a single mutation rule type, and having identified the three major controller operational stages (analysis, planning, execution), we must execute the tests with the controller in each of these stages, as it allows us to cover more cases and potentially disclose more robustness problems. Therefore, in each test, we must drive the system from an initial state to a target state by submitting the system to a changeload for a given amount of time (t1 in Figure 3). This target state is the one in which the system should be in order to start testing, and can correspond to any entry point to any of the three controller stages previously mentioned. With the controller in the target state, we can start applying the changes (of the same type) during a t2 amount of time (see Figure 3) and while the controller is on the target state. This period of time should be set to the typical time required to transition from the target controller state for the test to the next state. After this probe

Table 1 Mutation rules for probes

Type	Rule Name	Description
A. Message	1. MsgNull	Replace by null value
	2. MsgEmpty	Replace by empty string
	3. MsgPredefined	Replace by predefined string
	4. MsgNonPrintable	Replace by string with non-printable characters
	5. MsgAddNonPrintable	Add non-printable characters to the string
	6. MsgOverflow	Add characters to overflow max string size
B. Timestamp	1. TSEmpty	Replace by empty timestamp
	2. TSRemove	Remove timestamp from response
	3. TSInvalidFormat	Replace by timestamp with invalid format
	4. TSDateMaxRange	Replace date in timestamp by maximum valid
	5. TSDateMinRange	Replace date in timestamp by minimum valid
	6. TSDateMaxRangePlusOne	Replace date in timestamp by maximum valid plus one
	7. TSDateMinRangeMinusOne	Replace date in timestamp by minimum valid minus one
	8. TSDateAdd100	Add 100 years to date in timestamp
	9. TSDateSubtract100	Subtract 100 years from date in timestamp
	10. TSInvalidDate	Replace date in timestamp by invalid date (*e.g.*, 2/29/1985)
C. Var. Name	1. VNRemove	Remove variable name
	2. VNSwap	Replace by different valid variable name of same type
	3. VNSwapType	Replace by different valid variable name of different type
	4. VNInvalidFormat	Replace by variable name with invalid format
	5. VNNotExist	Replace by non-existing variable name
D. Var. Value	1. VVRemove	Remove variable value
	2. VVInvalidFormat	Replace value by one with invalid format
	Number	
	3. VVNumAbsoluteMinusOne	Replace by -1
	4. VVNumAbsoluteOne	Replace by 1
	5. VVNumAbsoluteZero	Replace by 0
	6. VVNumAddOne	Add 1
	7. VVNumSubtractOne	Subtract 1
	8. VVNumMax	Replace by maximum number valid for type
	9. VVNumMin	Replace by minimum number valid for type
	10. VVNumMaxPlusOne	Replace by maximum number valid for type plus one
	11. VVNumMinMinusOne	Replace by minimum number valid for type minus one
	12. VVNumMaxRange	Replace by maximum number valid for variable
	13. VVNumMinRange	Replace by minimum number valid for variable
	14. VVNumMaxRangePlusOne	Replace by maximum number valid for variable plus one
	15. VVNumMinRangeMinusOne	Replace by minimum number valid for variable minus one
	Boolean	
	16. VVBoolPredefined	Replace by predefined value

Table 2 Probe categories

Probe Usage Category	Controller Stage	Input Usage	Example Rainbow/Znn.com
Analysis	The controller analyzes the current state of the target system for detecting anomalies, and triggering adaptation if needed.	Anomaly detection.	Rainbow checks whether the current response time (through response time probes) in Znn.com is above the maximum acceptable response time threshold.
Planning	The controller determines if any adaptation plans can be applied to the system, and selects the best alternative.	Adaptation plan selection.	If the maximum response time is above threshold, Rainbow detects anomaly and determines the best adaptation strategy (based on response time and server fidelity probes).
Execution	The controller executes the selected course of action.	Control action selection.	Rainbow executes the selected adaptation strategy for reducing response time (monitors response time, server fidelity, and server load probes).

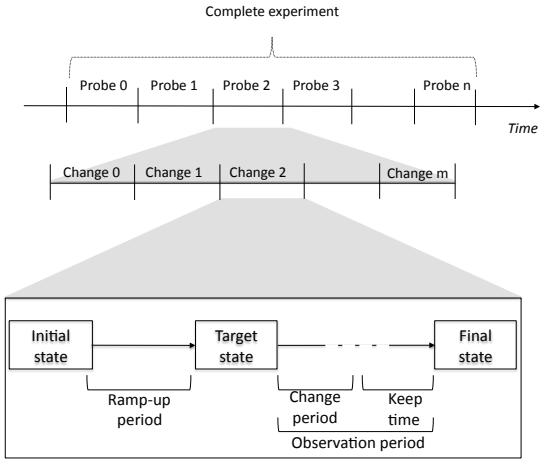

Fig. 3 Robustness testing procedure

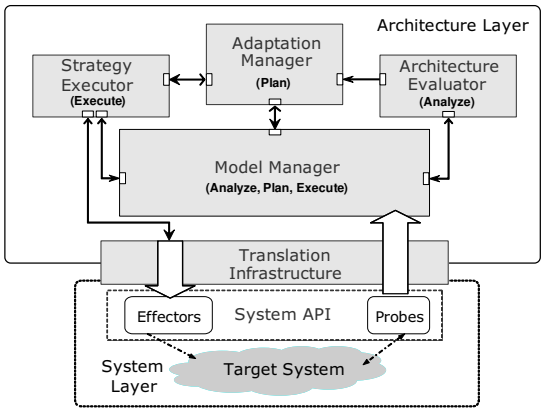

Fig. 4 The Rainbow framework

mutation period there is a t3 period which is the time required for the system to reach a final state, which marks the end of the current test, and corresponds to the completion of the controller's execution stage. At most, t3 should be set to the worst case execution duration found in the adaptation strategies specification. The t4 period (composed by t3 and t2) is an observation period that can be used to register any deviations from expected controller behaviour.

5 Experimental Evaluation

The aim of our experiments is assessing the validity of our approach to evaluate controller robustness in self-adaptive systems. In particular, we evaluate the robustness of Rainbow's controller (*i.e., Rainbow master*) on an implementation of the Znn.com case study described in Section 3.

5.1 The Rainbow Framework

In this paper, we focus on Rainbow [10], an architecture-based platform for self-adaptation, which provides a substantial base of reusable infrastructure through customization, which aims to reduce the cost of self-adaptive system development. Rainbow has distinctive features: an explicit architecture model of the target system, a collection of adaptation strategies, and utility preferences to guide adaptation.

The framework defined by Rainbow includes mechanisms for (Figure 4): monitoring a target system and its environment (using the observations for updating the architectural model of the target system), detecting opportunities for improving the system's quality of services (QoS), deciding the best course of adaptation based on the state of the system, and effecting the most appropriate changes.

Rainbow's component-and-connector architectural model of the target system is one of the main elements used in its decision-making process, using it to update monitored system information and reason about appropriate adaptation mechanisms for a particular situation.

The main components of the framework are:

- **Architecture Evaluator:** evaluates the model upon update to ensure that the system is operating within an acceptable range. If the evaluator determines that the system is not operating within the accepted range, it triggers the adaptation.
- **Adaptation Manager:** chooses a suitable strategy based on current state of the system (reflected in the architectural model).
- **Strategy Executor:** executes the strategy chosen by the adaptation manager on the running system via system-level effectors.
- **Model Manager:** updates the architecture model using the information observed in the system via probes.

5.2 Experimental Setup

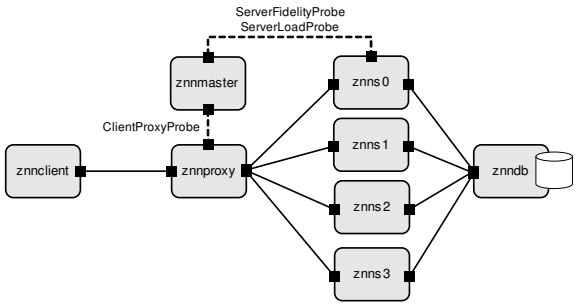

Fig. 5 Znn.com experimental setup

For our experimental setup, we deployed Rainbow and the corresponding implementation of Znn.com across seven different machines (Figure 5): znns0-3 are the four content servers running Apache v2.2.16, znndb is a common backend

database running mySql v14.14d5.1.61, from which the different servers extract the contents, and znnproxy is the proxy machine that runs the load balancing software (Apache running mod_proxy_balancer v2.1). The controller is deployed in a separate machine (znnmaster). All machines run Debian Linux v6.0.4, and have 512MB of memory. Moreover, an additional machine znnclient running JMeter v2.5.1 generates the traffic during the execution of the system.

To build the changeload used for our experiments, we identified the:

1. Workload and operating conditions for our change scenarios, which is characteristic of a slashdot-type effect, based on a sample collected by Juric [12], previously used for a general evaluation of the effectiveness of Rainbow in Znn.com [8]. In this case, scenarios have been scaled down to a duration of 5 minutes, which is enough to drive the controller through its different operational stages and apply the robustness tests.
2. Sets of probes for the different controller operational stages. The three last columns of Table 3 indicate the set of probes used during the analysis, planning, and execution stages of the controller. This information was identified by inspecting the specification of architecture models and adaptation strategies. Specifically, the use of a probe during the analysis stage can be determined by checking whether the constraints specified for the architecture model are defined over variables updated by a given probe. Moreover, an analogous process can be followed to identify probes used during the planning stage, which update information in variables used to specify applicability conditions of adaptation strategies. Finally, probes used during execution are identified by inspecting the predicates included in the code of the adaptation strategy itself.
3. Set of changes to be applied on our set of probes. Table 3 also indicates the set of non-applicable mutation rules to each of the probes, which are determined by the type of probe implemented, as well as, the data type and value range of the variables they update. Regarding probe implementation type, probes ServerLoadProbeT and ServerFidelityProbeT are implemented in Perl, whereas the ClientProxyProbeT is implemented in Java. In both cases, the length of strings is unrestrained, therefore mutation rule MsgOverflow (A6) is not applicable to any of the probes. In the particular case of Perl probes, the null datatype does not exist, disallowing the applicability of mutation MsgNull (A1). Regarding data types, all of the studied probes update numerical variables, disallowing the applicability of mutation rule VVBoolPredefined (D16). The only exception is the ServerLoadProbeT, which is not associated with a simple datatype and reports a message with a custom format in the variable value, therefore preventing the use of mutation rules D3-D16. Finally, the variables updated by some of the probes do not have a value range explicitly defined. In the case of probe ClientProxyProbeT there is an implicit lower bound of zero, due to the semantics of the information contained in the variable (*e.g.*, negative times would make no sense), but there is no upper bound, discarding the use of rules D12 and D14 that involve the maximum value range.

5.3 Experimental Results and Discussion

Each change scenario of the changeload results from combining the workload and operating conditions with a single change type based on an applicable mutation rule. In our changeload, each mutation rule gives way to up to three change scenarios (*i.e.,* applied during the analysis, planning, and execution stages, respectively), which are triggered in the time instant in which the controller enters the corresponding stage, and their duration is permanent. Overall, we run 209 robustness tests using our experimental setup (33 $\times$ 3 applicable mutation rules for ClientProxyProbeT, 21 $\times$ 2 applicable mutation rules for ServerLoadT, 34 $\times$ 2 applicable mutation rules for ServerFidelityProbeT,).

Table 4 details the experimental results obtained from the tests that apply the change scenarios based on each of the identified applicable mutation rules at each one of the controller stages. To begin with, 108 out of the 209 conducted tests uncovered robustness issues (51.6%). Moreover, one of the first observations that can be made is that no catastrophic, restart, nor hindering failures were identified during the tests. Although no catastrophic, restart, or hindering failures have been identified during our tests, these failure modes are still needed, as they portray relevant behaviours of the controller. Specifically, only 2.7% of the issues uncovered correspond to abort failures, which only occur on tests based on the mutation MsgNull (in this case, in the ClientProxyProbeT probe type, which is the only one implemented in Java). Specifically, this abort case consists of the same unhandled java.lang.NullPointerException in each of the three stages of the controller during the parsing of probe response with a regular expression matcher. It is worth mentioning that additional unhandled exceptions have been detected during the course of the experiments. However, these have not been considered in the results table, since they have been originated outside of the controller (concretely, on the response time probe itself).

Silent failures are by far the most frequent failure type discovered during the tests (97.3%). These mostly correspond to incorrect updates (or the lack thereof) of property values in the architecture model of the target system which are not acknowledged by the controller. In the case of the probes implemented in Perl (ServerLoadProbeT and ServerFidelityProbeT), when incorrect input is received by the con-

Table 3 Probe use per controller stage and applicable mutation rules for Znn.com

Probe Type	Description	Non-appicable mutation rules	Analysis	Planning	Execution
ClientProxyProbeT	Measures experienced response time in proxy.	A6,D12,D14,D16	x	x	x
ServerLoadProbeT	Measures the load of a given server.	A1,A6,D3-16	x		x
ServerFidelityProbeT	Reports the fidelity level of the contents served from a given server.	A1,A6,D16		x	x

Table 4 Robustness issues uncovered by the experiments

	Failures (A=Abort, S=Silent)															
	Analysis				Planning				Execution							
	ClientProxyProbeT		ServerLoadProbeT		ClientProxyProbeT		ServerFidelityProbeT		ClientProxyProbeT		ServerFidelityProbeT		ServerLoadProbeT	
Mutation Rule	A	S	A	S	A	S	A	S	A	S	A	S	A	S
MsgNull	1	1			1	1			1	1		1		
MsgEmpty		1		1		1		1		1		1		1
MsgPredefined		1		1		1		1		1		1		1
MsgNonPrintable		1		1		1		1		1		1		1
MsgAddNonPrintable		1		1		1		1		1		1		1
TSEmpty		1		1		1		1		1		1		1
TSRemove		1		1		1		1		1		1		1
VNRemove		1		1		1		1		1		1		1
VNSwap				1				1				1		1
VNInvalidFormat				1				1				1		1
VNNotExist				1				1				1		1
VVRemove		1		1		1		1		1		1		1
VVInvalidFormat		1		1		1		1		1		1		1
VVNumAbsoluteMinusOne		1				1		1		1		1		
VVNumMax		1				1				1				
VVNumMin		1				1		1		1		1		
VVNumMaxPlusOne		1				1		1		1		1		
VVNumMinMinusOne		1				1		1		1		1		
VVNumMinRangeMinusOne		1				1		1		1		1		
TOTAL/PROBE	1	16	0	12	1	16	0	18	1	16	0	18	0	12
TOTAL/STAGE	**A: 1, S: 28**				**A: 1, S: 34**				**A: 1, S: 46**					

troller, the update is ignored in all cases, and the property in the model is not updated. In contrast, in the Java probe (ClientProxyProbeT), properties are updated with clearly incorrect values (such as negative values in the case of ClientProxyProbeT with mutations VVNumAbsoluteMinusOne or VVNumMin), or not updated in some other cases (*e.g.*, mutations MsgNonPrintable or VNRemove).

As it can be observed, mutations that pertain the overall probe response message and the variable value (first and fourth group in Table 4, respectively) present the highest concentration of silent failures. In contrast, mutations that concern timestamps and variable names present silent failures only in cases in which the concrete element is removed (mutations TSEmpty, TSRemove and VNRemove). This is a consequence of the way in which the Rainbow master processes inputs from the probes. Messages sent from the probes are parsed in such a way that only the presence of a variable name and a timestamp in the message is assessed, but their concrete values are not checked syntactically nor semantically. However, this does not prevent the correct update of values in the architectural model of the system inside of the controller, which uses a unique probe identifier to update the value in the correct place in spite of incorrect variable names or timestamps in probe input.

In spite of the similarity of failure patterns across probes, we have been able to observe that there are slight differences in them directly related with their type of implementation: *(i)* all instances of abort failures are given when mutating the Java probe, and *(ii)* silent failures when mutating Perl probes always stops the updates of property values in the architecture model, in contrast with the Java probe, in which incorrect updates of values in the architectural model can also appear.

It is also worth mentioning that in spite of the similar failure patterns for the same probe across different controller stages, the specific failure instances discovered in the different controller stages are different. An instance of this is the mutation of the Java probe with the MsgNull, which results in the properties of the architecture model being updated with null values in tests conducted during the analysis stage, whereas in the planning and execution stages the last valid value on the model becomes frozen when the mutation rule is applied on the probe, and this can lead to completely different effects when considering the ensemble controller plus system.

Summarizing, although in general terms Rainbow master is fairly robust, experimental results have shown that our approach has been able to uncover a relevant set of robustness issues in the controller. Although in this particular case the identified pattern of robustness issues at the different stages of the controller differs only to a limited extent, this can be attributed to the particular architecture of Rainbow, which uses its model manager as a safeguard for the logic in the rest of the components used throughout the different operational stages. Moreover, the obtained results align with previous research, which has shown that robustness testing may disclose a small number of different issues, despite of their potentially high relevancy to the particular system being tested [18].

6 Conclusions

In this paper, we have presented a novel approach for testing the robustness of controllers for self-adaptive software systems. The approach consists in mutating the inputs provided by probes to the controller, according to a set of mutation rules and a target system's changeload, and taking into account the stateful nature of the controller. The proposal also includes an experimental setup and testing procedure specific to self-adaptive systems, as well as an adapted version of the CRASH failure scale that characterizes the different failure modes of a controller for self-adaptive software systems. We have evaluated the feasibility of our approach using Rainbow as a controller, which is based on an architecture-based self-adaptation framework, and in the context of the Znn.com case study, which reproduces the typical infrastructure for a news website.

Our experimental results have shown that the proposed approach has been able to discover a relevant number of controller failures that might impact negatively on the resilience of the self-adaptive system. However, despite the relevant number of failures uncovered, our approach has been unable to identify any catastrophic, restart, or hindering failures in the controller. Although this might be related to the restricted observability of the controller's internal behavior, other factors such as the architectural robustness of the controller might be a plausible explanation for such results. Indeed, the obtained results align with previous research on robustness testing, which has shown that these techniques may disclose a narrow range of different issues, despite of their potential relevance to the particular system being tested [18]. Regarding discovered failures, most of them correspond to silent ones and are distributed in similar patterns across the different probes and controller operational stages. However, it is worth observing that even if the failure categories coincide, the specific issues discovered are different between probes implemented with different technology, *i.e.*, Java and Perl. This is also true in some specific cases in which mutations on the same probe in different operational stages of the controller result in different kinds of silent failures.

Concerning future work, there are different lines of research that we intend to exploit based on the groundwork setup by this paper:

- Employ different controllers and additional case studies for assessing our approach in terms of its efficiency in uncovering faults in the controller of a self-adaptive software system.
- While the focus of this paper was the evaluation of the controller, there is the need for considering the self-adaptive system in its entirety, and this would inevitably lead to new challenges, such as, the necessity to consider the full state of the target system when evaluating the robustness of the entire system, *i.e.*, the controller plus the target system.
- Develop a framework for resilience evaluation of self-adaptive software systems based on our technique for evaluating the robustness of controllers. This work will be based upon previous work conducted on resilience evaluation of self-adaptive software systems [4,6], and will enable us to explore how robustness issues in the controller can influence the resilience of the overall self-adaptive system.
- Extend our robustness evaluation approach into the internal components of the controller that implement the MAPE-K loop. The idea is to test the interfaces between its components, in contrast with just focusing on the interface between the controller and the target system.
  A long term goal is to perform the type of evaluation described in this paper at run-time rather than development-time since the structure of a self-adaptive software system is expected to evolve during run-time.

Acknowledgements Co-financed by the Foundation for Science and Technology via project CMU-PT/ELE/0030/2009 and by FEDER via the "Programa Operacional Factores de Competitividade" of QREN with COMPETE ref.: FCOMP-01-0124-FEDER-012983.

References

1. Almeida, R., Vieira, M.: Benchmarking the resilience of self-adaptive software systems: perspectives and challenges. In: Proceedings of the 6th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS '11, pp. 190–195. ACM, New York, NY, USA (2011). DOI 10.1145/1988008.1988035
2. Andersson, J., Lemos, R., Malek, S., Weyns, D.: Software engineering for self-adaptive systems. chap. Modeling Dimensions of Self-Adaptive Software Systems, pp. 27–47. Springer-Verlag (2009)
3. Brun, Y., et al.: Software engineering for self-adaptive systems. chap. Engineering self-adaptive systems through feedback loops, pp. 48–70. Springer-Verlag (2009)
4. Cámara, J., de Lemos, R.: Evaluation of Resilience in Self-Adaptive Systems Using Probabilistic Model-Checking. In: 7th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS 2012), pp. 53–62. IEEE (2012)
5. Cámara, J., de Lemos, R., Laranjeiro, N., Ventura, R., Vieira, M.: Robustness Evaluation of Controllers in Self-Adaptive Software Systems. In: 6th Latin American Symposium on Dependable Computing (LADC 2013), pp. 411–420. IEEE (2013)
6. Cámara, J., de Lemos, R., Vieira, M., Almeida, R., Ventura, R.: Architecture-Based Resilience Evaluation for Self-Adaptive Systems. Computing (2013). Available online. DOI:10.1007/s00607-013-0311-7
7. Cheng, B.H., et al.: Software engineering for self-adaptive systems. chap. Software Engineering for Self-Adaptive Systems: A Research Roadmap, pp. 1–26. Springer-Verlag (2009)
8. Cheng, S.W., Garlan, D., Schmerl, B.R.: Evaluating the Effectiveness of the Rainbow Self-Adaptive System. In: 4th International Workshop on Software Engineering for Adaptive and Self-Managing Systems (SEAMS 2009), pp. 132–141. IEEE (2009)
9. Cotroneo, D., Di Leo, D., Natella, R., Pietrantuono, R.: A case study on state-based robustness testing of an operating system for the avionic domain. In: Computer Safety, Reliability, and Security, *Lecture Notes in Computer Science*, vol. 6894, pp. 213–227. Springer Berlin / Heidelberg (2011)
10. Garlan, D., Cheng, S.W., Huang, A.C., Schmerl, B.R., Steenkiste, P.: Rainbow: Architecture-Based Self-Adaptation with Reusable Infrastructure. IEEE Computer **37**(10), 46–54 (2004)
11. Gray, J.: Benchmark Handbook: For Database and Transaction Processing Systems. Morgan Kaufmann Publishers Inc. (1992)
12. Juric, M.: Slashdotting of mjuric/universe (2004). `http://www.astro.princeton.edu/~mjuric/universe/slashdotting/`
13. Kanoun, K., Spainhower, L.: Dependability Benchmarking for Computer Systems. Wiley-IEEE Computer Society Pr (2008)
14. Kephart, J.O., Chess, D.M.: The vision of autonomic computing. Computer **36**, 41–50 (2003)
15. Koopman, P., DeVale, J.: Comparing the robustness of POSIX operating systems. In: Twenty-Ninth Annual International Symposium on Fault-Tolerant Computing, pp. 30–37 (1999)
16. Koopman, P., DeVale, J.: Comparing the robustness of posix operating systems. In: Proceedings of the Twenty-Ninth Annual International Symposium on Fault-Tolerant Computing, FTCS '99, pp. 30–. IEEE Computer Society (1999)
17. Laprie, J.C.: From Dependability to Resilience. In: DSN Fast Abstracts. IEEE Computer Society (2008)
18. Laranjeiro, N., Vieira, M., Madeira, H.: Experimental robustness evaluation of JMS middleware. In: IEEE International Conference on Services Computing (SCC 2008), pp. 119–126. IEEE Computer Society (2008)
19. de Lemos, R., et al.: Software Engineering for Self-Adaptive Systems: A Second Research Roadmap. In: Software Engineering for Self-Adaptive Systems 2, no. 7475 (to appear) in Lecture Notes in Computer Science. Springer-Verlag (2012)
20. Madeira, H.: Towards a security benchmark for database management systems. In: Proceedings of the 2005 International Conference on Dependable Systems and Networks, DSN '05, pp. 592–601. IEEE Computer Society (2005)
21. Micskei, Z., Majzik, I., Tam, F.: Robustness testing techniques for high availability middleware solutions. In: in Proc. of Int. Workshop on Engineering of Fault Tolerant Systems (2006)
22. Mukherjee, A., Siewiorek, D.: Measuring software dependability by robustness benchmarking. Transactions on Software Engineering **23**(6), 366–378 (1997)
23. Oreizy, P., Gorlick, M.M., Taylor, R.N., Heimbigner, D., Johnson, G., Medvidovic, N., Quilici, A., Rosenblum, D.S., Wolf, A.L.: An architecture-based approach to self-adaptive software. IEEE Intelligent Systems **14**, 54–62 (1999)
24. Pan, J., Koopman, P., Siewiorek, D.P., Huang, Y., Gruber, R., Jiang, M.L.: Robustness testing and hardening of CORBA ORB implementations. In: The 2001 International Conference on Dependable Systems and Networks (DSN 2001), pp. 141–150. IEEE Computer Society (2001)
25. Rodríguez, M., Salles, F., Fabre, J.C., Arlat, J.: MAFALDA: microkernel assessment by fault injection and design aid. In: The Third European Dependable Computing Conference on Dependable Computing, pp. 143–160. Springer-Verlag (1999)
26. Rodriguez, M., Salles, F., Fabre, J.C., Arlat, J.: Mafalda: Microkernel assessment by fault injection and design aid. In: Proceedings of the Third European Dependable Computing Conference on Dependable Computing, EDCC-3, pp. 143–160. Springer-Verlag (1999)
27. Salva, S., Rabhi, I.: Stateful web service robustness. In: Fifth International Conference on Internet and Web Applications and Services. Iaria (2010)
28. Shelton, C., Koopman, P., Devale, K.: Robustness testing of the microsoft win32 API. In: International Conference on Dependable Systems and Networks (DSN 2000), pp. 261–270 (2000)
29. Vieira, M., Laranjeiro, N., Madeira, H.: Benchmarking the robustness of web services. In: 13th IEEE Pacific Rim Dependable Computing Conference (PRDC 2007), pp. 322–329. IEEE (2007)