

On Comparing and Complementing Two MBT Approaches

Maximiliano Cristiá
Flowgate and CIFASIS
Rosario, Argentina
mcrastia@flowgate.net

Valdivino Santiago
INPE
São José dos Campos, Brazil
valdivino@das.inpe.br

N. L. Vijaykumar
INPE
São José dos Campos, Brazil
vijay@lac.inpe.br

Abstract—At INPE¹ researchers and software engineers have been using Statechart-based testing for some time to test on-board satellite software. On the other hand, a group of researchers at CIFASIS² and Flowgate Consulting have been applying Z-based testing for unit testing. Both groups started to compare their approaches and tools a year ago. What started as a comparison to share ideas and results, is now turning into the realization that actually both techniques complement and benefit from each other, yielding a more effective and wider Model-Based Testing (MBT) approach. In this paper we present the results obtained so far in comparing and complementing these two MBT techniques.

I. INTRODUCTION

Satellite software is a mission critical, reactive and data handler program that must be validated and verified thoroughly. In these cases, verification and validation always comprises testing although other techniques can be applied too. Testing a program which presents a reactive behavior and data handling functions as well, is a complex task that does not seem to be extensively addressed in an industrial setting. Furthermore, if we also want Model-Based Testing (MBT) to test both units and the complete system with tool support, then the informed results rapidly approaches to zero. In [1] and [2] we could not find such issues being addressed, although [3] deals with them. Hence, in this paper we address the problem outlined above, in the context of space application software.

As we indicate in the paper's title, this work is the outcome of what started as a comparison between two MBT approaches and later became a combination of both techniques. We started our joint project with the intention to see how we could help INPE's test designers to better and efficiently test satellites' on-board software. At the beginning only one thing was clear: both groups were working on MBT but with different notations and tools. Then, we decided to demonstrate our methods to each other using the same real-world case study proposed by INPE. The first conclusion was that both methods were functionally testing the same system but in different ways. Latter we extended Paradkar's work [21] in order to accomplish the comparison between our approaches. Finally, we realized that our techniques could be combined into a

single approach addressing model-based test case generation with the main benefits of both techniques: Statecharts to model behavior and Z to model the data space.

In Sections II and III we briefly introduce the method used by each group. Section IV shows the results of comparing our methods. Perhaps the main contribution of this paper lays in Section V where we suggest a proposal for combining Statechart-based and Z-based testing. Sections VI and VII compare our work with similar approaches and present our conclusions, respectively.

II. STATECHARTS AND FINITE STATE MACHINES FOR TESTING AT INPE

INPE has been using MBT within research projects for system and acceptance testing. Statecharts [4], [5] and Finite State Machines (FSMs) [6] are the main techniques used for modeling the behavior of the Implementation Under Test (IUT) for testing purposes. The application domain is software embedded into on-board computers of scientific satellites and balloons under development at INPE. These MBT approaches have shown efficiency in detecting defects in the source code.

One of the main advantages of FSMs is simplicity. Reactive systems, protocol implementations, classes of Object-Oriented applications are some examples commonly address-ed by FSM modeling. Several test criteria (methods) may be used for model-based test case generation regarding the FSM technique. A few of such methods are the Transition Tour (TT), Distinguishing Sequence (DS), Unique Input/Output (UIO), W [7], switch cover (1-switch) [8] and state counting [9]. Despite their adoption by many researchers, FSMs are not too adequate for representing features such as parallelism and hierarchy. On the other hand, Statecharts provide a simple way to represent these characteristics. There are several approaches proposed to generate test cases from Statecharts models [10] [4].

System and acceptance testing are applied considering the entire software product. If a professional wants to develop model-based test case generation relying on, for example, FSMs or Statecharts, he/she must develop the state-transition diagrams. Due to the state explosion problem, a test designer at INPE usually breaks down the entire system based on usage scenarios. Models are then derived to address each scenario and, provided they are small enough, test cases are generated from them.

¹INPE – National Institute for Space Research.

²CIFASIS – French Argentine International Center for Information Systems and Sciences.

INPE has been developing an environment named *Geração Automática de Casos de Teste Baseada em Statecharts* (GTSC) that allows test designers to model software behavior using Statecharts and/or FSMs in order to automatically generate test cases based on some test criteria for FSM and some for Statecharts [5]. At present, GTSC implements a version of switch cover, UIO and DS test criteria for FSM models and two test criteria from the *Statechart Coverage Criteria Family* (SCCF) [11] for Statecharts models: all-transitions and all-simple-paths.

GTSC transforms a Statecharts model into a flat FSM, i.e. a model where all hierarchical and orthogonal features of the Statecharts were removed. Each state of the resulting flat FSM is actually a configuration of active BASIC states of the input model at a certain instant of time. This flat FSM is indeed the basis for test case generation. Hence, a test designer may follow two approaches:

- 1) If a SCCF test criterion will derive test cases, GTSC must adapt the flat FSM to resemble a reachability tree [12]. Thus, based on the selected test criterion of SCCF and on this tree, test cases are created.
- 2) If an FSM test criterion is the option, it is only necessary for the user to choose among the available criteria for FSM and instruct the environment to generate test cases, based on the flat FSM.

GTSC was able to generate flat FSMs with as many as 40 states (configurations) and more than 300 transitions, and test suites with up to 265 optimized test cases, showing its potential scalability for handling complex systems [5]. INPE has also been developing a Web tool that allows model-based test case generation named WEB-Perform Charts [30]. WEB-PerformCharts works similar to GTSC, by transforming a Statecharts model into a flat FSM, and it has implemented the TT, switch cover and UIO test criteria for FSM models. Such a tool addresses collaborative work where different teams cooperate with an objective of reaching a specific goal, and it has generated flat FSMs as complex as the ones created by GTSC.

III. Z-BASED TESTING WITH FASTEST

Phil Stocks and David Carrington introduced in [13], [14], [15] the Test Template Framework (TTF) to conduct MBT of Z specifications [16]. TTF includes a rigorous and disciplined technique for defining and structuring abstract test templates and cases³. They also proposed new testing tactics particularly well suited to the Z notation. Testing tactics are the mechanisms used to partition the input space into test templates and, in turn, test templates into more test templates, thus building a so called *testing tree*. Test cases are elements selected from the leaves of the testing tree.

Fastest [17] is a flexible, efficient and automatic implementation of the TTF developed conjointly by CIFASIS and Flowgate Consulting. Currently, Fastest automates test suite definition and test case derivation for unit testing. Fastest

receives a Z specification in \LaTeX format using the CZT package [18], [19]. Then, the user has to enter a list of the operations to test, as well as the tactics to apply to each of them. In a third step Fastest automatically generates the testing tree of each operation. After the trees are generated, the user can browse them and their test classes, and he/she can prune any node, both manually or automatically. Once the user is done with pruning, he/she can instruct Fastest to find one abstract test case for each leaf in all the test trees. Although the method to find abstract test cases has proved to be quite automatic, it is worth to say that it does not guarantee to find abstract test cases for all test objectives. In those cases, the engineer can help Fastest to find a test case by issuing a rather straightforward command. The user can export all the results –testing trees, test classes and abstract test cases– in \LaTeX format.

Fastest was envisioned as a client-server application. The main reason for thinking in a distributed system came from the realization that calculating abstract test cases from test objectives in large projects could be a hard computing problem, but highly parallelizable as well. Then, a scalable application using the idle computer power present in a corporate network, became an appealing option. However, in such a large project there is shared information –such as the definition or parametrization of some testing tactics, test cases already calculated, theorems that help to prune testing trees, etc.– that all the clients and servers should be able to access. Hence, a typical Fastest installation has a data server that is known to all other processes, some client processes and a number of testing servers.

IV. COMPARISON: CASE STUDIES

As we pointed out earlier, this paper is the result of a joint effort between two R+D institutions from Argentina and Brazil, who belong to the REVVIS project funded by CYTED. The cooperation started by comparing the MBT techniques used at each institution. The second step was to define a common problem to work with so both groups had a common workbench. The first problem was proposed by INPE's group. CIFASIS researchers then developed a Z model and applied Fastest in order to generate test cases. INPE researchers had already a Statecharts model and the test cases generated by the GTSC environment. The comparison was then extended to two more examples: one proposed by INPE and the other proposed by CIFASIS.

We followed the same methodology for all the problems. First, the party proposing the problem delivers an informal, natural language requirements specification. Second, the other party starts to write a formal model. Third, if the party writing the model finds some problem (incompleteness, inconsistency) or misunderstands the requirements, the other party will send corrections/explanations. Fourth, when the model has been finished the party who wrote it will apply its MBT methodology and tool.

In the next subsections we briefly describe the three problems.

³Test templates can also be called *test suites*, *test classes* or *test objectives*.

A. EXP – OBDH Communication Protocol

The first problem proposed by INPE was the software that will be embedded into an astrophysical experiment computer, hereafter called EXP, of a Brazilian scientific satellite. A proprietary protocol was specified for the communication between EXP and the On-Board Data Handling (OBDH) computer. OBDH is the satellite platform computer to process platform and payload information and to generate and format data that has to be transmitted to Ground Stations.

OBDH sends one out of nine commands at a time to EXP, which returns an answer to OBDH. Each command must arrive within certain time constraints. Commands ask EXP to perform some operations or to return some data about its state. There are simple commands and there are more complex commands to transmit scientific data acquired by the payload, to dump EXP's memory, to load data sent by OBDH, etc.

B. A Simple Scheduler

This problem was proposed by CIFASIS borrowing it from [20]. The problem is about the basic operations of a simple scheduler. The environment can buffer processes for latter execution and can withdraw a process from the waiting list. On the other hand, the scheduler can swap between the active process and one other process ready for execution. The idea behind this problem was twofold: (a) to propose a simpler problem, and (b) to apply our techniques to a somewhat less reactive system.

C. SWPDC

The third problem was a software product specified and developed in the scope of the *Qualidade do Software Embarcado em Aplicações Espaciais* (QSEE) research project at INPE [31]. SWPDC is the software embedded into the Payload Data Handling Computer (PDC) and it is the hardest problem of all we have analyzed. Although similar in conception to the case study described in Section IV-A, SWPDC is much more complex because it handles not only scientific and dump data (e.g. housekeeping data), but also accomplishes data memory management, implements flow control mechanisms, etc. Data transmission is more complex since SWPDC has to keep record of the last transmitted frame because OBDH can ask it again if some problem during transmission was detected.

D. Summary

After applying each technique to the preceding problems, the first and quite obvious conclusion was that all looked very different: models, test cases and methods or tactics were, initially, hard to compare. Hence, we needed some framework inside which we can compare our techniques and tools. We, then, searched for articles comparing MBT techniques and tools [21], [1], [22]. Eventually, we followed [21] and extended the set of dimensions defined by the author to base the comparison. We divided these dimensions into two groups summarized in Tables I and II. Table I includes those dimensions that are, at least to some extent, independent of the

cases studies, while Table II reunites those dimensions whose values depend on each case study.

The dimensions named **Notation concepts** and **MBT concepts** include the new concepts a software engineer with a basic knowledge on formal modelling has to learn in order to apply each technique. The dimensions whose values are marked with **X** indicate which technique is more suitable for each dimension –this does not necessarily mean that the other technique cannot be applied at all. In column **Number of test cases** of Table II, we show for GTSC the number of test cases and (*l*) the number of test steps since “test step” is a meaningful concept regarding FSM-based MBT techniques. On the other hand, for Fastest we just give the number of test cases since each test case has always only one test step. The dimension called **Ratio** is the ratio between the two left rows. By indicating the **Model size** we tried to give a broad idea of the model complexity. For the Z models we give the number of Z- \LaTeX lines of code, while for the Statecharts models we give the number of states and (*l*) the number of transitions of the flat FSM. Although these measures are incomparable to each other, they give an idea of the relative complexity: (a) between the case studies, and (b) with respect to other models that can be found in the literature.

Table I shows that both techniques are rather similar, although GTSC can be applied to more phases of the testing process. As the reader can easily note, Table II shows that GTSC outperforms Fastest when reactive systems are considered (EXP and SWPDC), but Fastest wins when our tools are applied to a (more) information oriented system (Scheduler).

Although this comparison was useful for us, we moved one step forward because we saw a possibility for complementing our techniques, as is shown in the next section.

V. A COMPLEMENTARY APPROACH

As we said in Table I test cases generated by GTSC are sequences of state transitions, while test cases generated by Fastest are bindings or assignments between state variables and constant values. The following sequences are two test cases generated by GTSC for the EXP case study.

$$0xEB \rightarrow 0x92 \rightarrow T1B \rightarrow Size38_E \rightarrow Data0x100_E \\ \rightarrow TimerTimeout$$
$$0xEB \rightarrow 0x92 \rightarrow T1B \rightarrow Size38_E \rightarrow Data0x100_E \\ \rightarrow TypeOkSizeOkDataOk_E \rightarrow Cks_Ok$$

They correspond to OBDH sending a memory load command to EXP –i.e. OBDH sends a program to EXP that it must load into its own memory. *0xEB* and *0x92* identify the beginning of a new order coming from OBDH. *T1B* means that the sequences correspond to a memory load command. *Size38_E* and *Data0x100_E* are the size in Bytes of a piece of program, and the address in memory where this piece of code will be loaded followed by the code, respectively. The event *TypeOkSizeOkDataOk_E* means the data (code) sent by OBDH was received correctly by EXP. Note that the first test sequence ends in a timeout event. This means

Dimension	GTSC – INPE	Fastest – CIFASIS & Flowgate
Notation concepts	Statecharts: XOR and AND states, external events, actions (internal events or outputs), conditions, hierarchy, parallelism and machine synchronization, broadcasting, shallow and deep history, PcML language	Z (subset): first order logic, state and operation schema, schema language, typed logic, typed set theory, before and after state convention, input and output variables convention, logic and mathematics for modelling
MBT concepts	reachability tree, state configuration, Statecharts flattening, determinism, test criteria for FSM (switch cover, UIO and DS) and for Statecharts (all-transitions and all-simple-paths)	testing tactics (standard partitions, disjunctive normal form, free types), testing tree, domain partition, valid input space
Computability issues	state explosion	huge finite models
Computability solutions	duplicate nodes (Statecharts test case generation)	parallelization, user assistance
Test objectives	usage scenarios and models to represent them	written as Z schema boxes
Test cases	sequences of events producing state transitions	bindings between state or input variables, and constant values
Unit testing		X
Integration testing	X	
System testing	X	
Acceptance testing	X	
Reactive systems	X	
Information systems		X

TABLE I
COMPARISON CRITERIA. CASE-STUDY-INDEPENDENT DIMENSIONS.

Case study	Dimensions							
	Computing time		Number of test cases		Ratio		Model size	
	GTSC	Fastest	GTSC	Fastest	GTSC	Fastest	GTSC	Fastest
EXP	0:51	124:00	78/436	112	0:0.6	1:06	17/63	608
Scheduler	3:20	3:00	5/43	29	0:40	0:06	7/16	240
SWPDC	11:42	158:00	30/663	117	0:23	1:21	512/522	1,238

TABLE II
COMPARISON CRITERIA. CASE-STUDY-DEPENDENT DIMENSIONS. TIMES ARE IN MINUTES.

that OBDH has taken too long in sending all the data, and so actually no memory load is performed. The second test sequence describes a successful command, i.e. EXP loads the code into its memory.

The following Z schema boxes are two test cases generated by Fastest for the same memory load command.

MemoryLoad_SP_15_TCASE

MemoryLoad_SP_15

dumping = no \wedge *mep* = 0 \wedge *ccmd* = *ML*
low? = *no* \wedge *ped* = 45 \wedge *ctime* = 45
mdp = 45 \wedge *data?* = \langle *mdata0*, *mdata1* \rangle
memp = \emptyset \wedge *acquiring* = *no* \wedge *waiting* = *no*
memd = \langle *mdata0* \rangle \wedge *waitsignal* = *no*
addr? = 1 \wedge *sending* = *no* \wedge *mode* = *COF*

MemoryLoad_SP_43_TCASE

MemoryLoad_SP_43

dumping = no \wedge *mep* = 2 \wedge *ccmd* = *RM*
low? = *no* \wedge *ped* = 2 \wedge *ctime* = 2 \wedge *mdp* = 2
data? = \emptyset \wedge *memp* = \emptyset \wedge *acquiring* = *no*
waiting = *no* \wedge *memd* = \langle *mdata0* \rangle
waitsignal = *no* \wedge *addr?* = 2
sending = *no* \wedge *mode* = *COF*

Observe that each test case is a set of assignments⁴. Each assignment sets a constant value for one of the state and input variables. Regarding this command, the most important variables are *data?*, which is the input variable representing the new program; *addr?*, representing the initial address where the program must be loaded; and *memd*, which is the state variable representing EXP's memory. Then, roughly speaking, the first test case says that OBDH has sent a two byte long program to be loaded starting at the first memory location while EXP's memory holds a one byte long program; and, on the other hand, the second test case says that OBDH has sent an empty program –that should be loaded after the program currently being held by EXP– while EXP's memory is occupied by a single byte program.

After arriving at this point, an approach that combines both techniques emerged quite easily. We consider all the test sequences generated by GTSC that include in the execution of a data intensive, underspecified⁵ operation –with respect to the Statecharts model– and at that point we execute all the test

⁴Actually they are not assignments but propositional equalities.

⁵Underspecification in Statecharts models comes into picture due to two main reasons. First, Statecharts are based on FSM, and Z on first order logic. First order logic is more expressive than FSM, regarding data transformation operations. Second, the ability (or lack of ability) of the test designer in adequately modeling the IUT for testing. And this lack of ability occurs probably because a test designer relies the development of the models based on documents (requirements specifications, ...) which present serious issues. In some cases, there is no documentation at all available which is a worse situation.

cases generated by Fastest for that operation. For example, since the first GTSC test sequence written above does not include in the execution of a data intensive operation, then we do not execute any Z test case for it; however, we do execute 28 Z test cases for the second test sequence since it represents EXP loading its memory with a new program. In this way, the software is tested as a whole and every time it gets to a point where a complex data-intensive operation has to be executed, then this operation is fully tested too.

Now we present the approach in a more formal and systematic way. Let E be the set of events in the Statecharts model and D the subset of E that represents the execution of some data-intensive operation, which have also been specified as Z operations. Let S be the set of test sequences derived from the Statecharts model. Now, consider the subset of S , S_D , including all the test sequences that contain some element of D . For each d in D , define T_d as the set of tests derived from the Z model for unit d ; each t in T_d is a call to d with some specific input parameters. If P is the program that has been modeled in Statecharts and Z, then test P with the set of test sequences defined as follows (\wedge is sequence concatenation):

$$S_T = \{s_i \wedge \langle t \rangle \wedge s_f \mid \\ (\exists d \in D : s_i \wedge \langle d \rangle \wedge s_f \in S_D \wedge t \in T_d)\} \\ \cup (S \setminus S_D)$$

In other words, we must break any test sequence in the points where there is a call to a data-intensive operation and replace these points with each test case generated from the Z model for the corresponding operation.

As this formalization shows, the method can be made automatic by simply using a convenient naming convention.

Coming back to the EXP-OBDDH case study, there are two data intensive operations: memory load and transmit data. Fastest yielded 28 and 20 test cases, respectively, for these operations. On the other hand, GTSC generates two test sequences including the execution of these operations. Then, the original 78 test cases generated by GTSC become 126, improving the original testing in about 60%. Furthermore, there is not only an improvement in the number of test cases but also (a) in the chances to uncover defects (faults) in complex operations that would not be thoroughly tested by any method alone; and (b) in the fact that the test cases provided by Fastest give the exact input constants that must be provided to the IUT, thus augmenting the proportion of automatic steps of the testing process.

The research methodology that we set before getting these results made the Z models unnecessarily complex because they describe not only those operations underspecified in the Statecharts model but also all the simpler and reactive ones. Now, we know that the Z model must describe only data intensive operations. In this way we can reduce both the effort and the size (complexity) of the Z model. For instance, the EXP-OBDDH Z model is 608 ZLOC (Table II) but only less than 241 are involved in the data intensive operations.

It is worth to be mentioned that Statecharts can be used to model data-intensive systems, and, on the other hand, Z can be

used to model reactive systems. Hence it would be possible to use only one approach, but we think it would be easier and more intuitive to combine both approaches as we have described in this section.

VI. RELATED WORK

Our method has an advantage over [23] because it is based on two industrial-strength formal notations, one of them being part of the UML specification [24], and for which exist commercial and open source tools.

An interesting discussion in [1] shows that the authors were not successful in comparing MBT with conventional testing techniques as they expected some strong studies while investigating MBT. However, the authors discuss some interesting aspects. For instance, reviews in the area of MBT showed that UML models stand out for object-oriented paradigm while for paradigm based on formal specification non-UML models stand out. Their study also hints that non-UML approaches were applied on safety-critical or embedded systems. They point out that in spite of MBT appeal, training and corresponding costs are to be taken into consideration.

Grieskamp et al. demonstrated the use of the agenda approach to support the systematic development of requirement specifications for a particular class of embedded safety-critical systems [25]. An agenda is a list of steps to be accomplished when carrying out some task in the context of software engineering. To this end, they used the μ SZ notation which provides a combination of the Statechart languages [26] (Activity Charts and Statecharts), Z, and an extension of Z by temporal interval logic [27]. Basically, they used Statechart's Activity Charts to represent process structure, Z to specify the data space of the process, and Statecharts to specify the dynamic behavior of the process. However, it is not clear whether agendas with their derived Z schemata and Statecharts models were developed to support Model-Based Testing. Hierons and others in [3] define a MBT technique for the μ SZ notation although they do not show a tool set. The main difference with our work is that we needed to develop a combined MBT methodology suitable for the tools each team already had. For instance, we had to combine a unit-testing oriented tool such as Fastest, with a system-testing oriented tool such as GTSC to get the best of them both.

In his MSc. thesis [28] Mäkinen studied on how an organization can make use of model based approach for software testing by showing the financial benefits in adopting such approach. It compares the traditional testing with the MBT approach and concludes that there are, in fact, financial benefits to organizations by adopting MBT. Moreover, the MBT approach saves time in test execution. The case study was a user interface. However, the author suggests investigating other complex model representation if this approach is to be used to other software engineering levels.

In [29], the authors briefly present the following models used in software testing: FSMs, Statecharts, UML and Markov chains. They provide a detailed definition of FSM and discuss some aspects of MBT such as the importance of understanding

the system to be tested and how to build a model besides generating tests and analyzing the results. They also point out some concerns to be considered on state explosion, coverage, oracle and automation.

VII. CONCLUSIONS AND FUTURE WORK

First, we have compared two MBT techniques and tools, and then we have shown how to combine those techniques to improve the testing of systems including both reactive and data intensive functions. The main conclusions of this article are: (a) it is rather easy to combine the approaches, and (b) that the combined approach could potentially uncover many defects that each technique alone cannot.

There is a lot of work to do in the future. First, we need to assess the amount of faults uncovered by the combined approach by applying, for instance, mutation analysis. In doing so, we can compare the combined methodology with respect to applying the individual methodologies developed so far. Second, we want to work on new case studies so we can apply what we have learned so far –for example, writing shorter Z models that efficiently complement the Statecharts model. Third, we want to integrate our tools so the whole process will be more automatic.

ACKNOWLEDGMENTS

We would like to thank CYTED for funding part of this work through action REVVIS (number 507AC0326).

REFERENCES

- [1] A. D. Neto, R. Subramanian, M. Vieira, G. H. Travassos, and F. Shull, "Improving evidence about software technologies: A look at model-based testing," *IEEE Softw.*, vol. 25, no. 3, pp. 10–13, 2008.
- [2] R. M. Hierons and et.al., "Using formal specifications to support testing," *ACM Comput. Surv.*, vol. 41, no. 2, pp. 1–76, 2009.
- [3] R. M. Hierons, S. Sadeghipour, and H. Singh, "Testing a system specified using statecharts and z," *Information and Software Technology*, vol. 43, no. 2, pp. 137–149, February 2001.
- [4] V. Santiago, A. S. M. do Amaral, N. L. Vijaykumar, M. d. F. Mattiello-Francisco, E. Martins, and O. C. Lopes, "A practical approach for automated test case generation using statecharts," in *COMPSAC '06*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 183–188.
- [5] V. Santiago, N. L. Vijaykumar, D. Guimarães, A. S. Amaral, and E. Ferreira, "An environment for automated test case generation from statechart-based and finite state machine-based behavioral models," in *ICSTW '08*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 63–72.
- [6] A. M. Ambrosio, M. de Fátima Mattiello-Francisco, V. A. Santiago, W. P. da Silva, and E. Martins, "Designing fault injection experiments using state-based model to test a space software," in *LADC*, 2007, pp. 170–178.
- [7] D. P. Sidhu and T.-k. Leung, "Formal methods for protocol testing: A detailed study," *IEEE Trans. Softw. Eng.*, vol. 15, no. 4, pp. 413–426, 1989.
- [8] S. Pimont and J.-C. Rault, "A software reliability assessment based on a structural and behavioral analysis of programs," in *ICSE '76*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1976, pp. 486–491.
- [9] A. Petrenko and N. Yevtushenko, "Testing from partial deterministic FSM specifications," *IEEE Trans. Comput.*, vol. 54, no. 9, pp. 1154–1165, 2005.
- [10] R. V. Binder, *Testing object-oriented systems: models, patterns, and tools*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.
- [11] S. R. S. Souza, "Validao de especificaes de sistemas reativos: Definio e anlise de critrios de teste," Ph.D. dissertation, Instituto de Física de São Carlos, Universidade de São Paulo, 2000.
- [12] P. C. Masiero, J. C. Maldonado, and I. G. Boaventura, "A reachability tree for statecharts and analysis of some properties," *Information and Software Technology*, vol. 36, no. 10, pp. 615–624, October 1994.
- [13] P. Stocks and D. Carrington, "A Framework for Specification-Based Testing," *IEEE Trans. on Soft. Eng.*, vol. 22, no. 11, pp. 777–793, Nov. 1996.
- [14] P. Stocks, "Applying formal methods to software testing," Ph.D. dissertation, Department of Computer Science, University of Queensland, 1993.
- [15] I. Maccoll and D. Carrington, "Extending the Test Template Framework," in *Proceedings of the Third Northern Formal Methods Workshop*, 1998.
- [16] J. M. Spivey, *The Z Notation: A Reference Manual*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989.
- [17] M. Cristiá and P. Rodríguez Monetti, "Implementing and applying the Stocks-Carrington framework for model-based testing," in *ICFEM*, ser. LNCS vol. 5885. Springer, 2009, pp. 167–185.
- [18] Community Z Tools, <http://czt.sourceforge.net>.
- [19] P. Malik and M. Utting, "CZT: A Framework for Z Tools," in *ZB. Lecture*. Springer, 2005, pp. 65–84.
- [20] M. Utting and B. Legeard, *Practical Model-Based Testing: A Tools Approach*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2006.
- [21] A. Paradkar, "Case studies on fault detection effectiveness of model based test generation techniques," in *A-MOST '05*. New York, NY, USA: ACM, 2005, pp. 1–7.
- [22] A. Sinha, C. E. Williams, and P. Santhanam, "A measurement framework for evaluating model-based test generation tools," *IBM Systems Journal*, vol. 45, no. 3, pp. 501–514, 2006.
- [23] A. J. Offut and S. Liu, "Generating test data from SOFL specifications," *J. Syst. Softw.*, vol. 49, no. 1, pp. 49–62, 1999.
- [24] Object Management Group, "Unified modeling language – resource page," <http://www.uml.org>.
- [25] W. Grieskamp, M. Heisel, and H. Drr, "Specifying embedded systems with Statecharts and Z: an agenda for cyclic software components," *Science of Computer Programming*, vol. 40, no. 1, pp. 31 – 57, 2001.
- [26] D. Harel, H. Lachover, A. Naamad, A. Pnueli, M. Politi, R. Sherman, A. Shull-Trauring, and M. Trakhtenbrot, "Statemate: A working environment for the development of complex reactive systems," *IEEE Trans. Soft. Eng.*, vol. 16, no. 4, pp. 403–414, 1990.
- [27] R. Büsow and W. Grieskamp, "Combining Z and temporal interval logics for the formalization of properties and behaviors of embedded systems," in *LNCS, ASIAN'97*, vol. 1345, 1997, pp. 46–56.
- [28] M. A. Mäkinen, "Model based approach to software testing," Master's thesis, Department of Electrical and Communications Engineering Networking Laboratory, Helsinki University of Technology, 2007.
- [29] I. El-Far and J. Whittaker, *Encyclopedia of Software Engineering*. John Wiley & Sons, 2001, ch. Model-Based Software Testing, pp. 825–837.
- [30] A. O. Arantes, N. L. Vijaykumar, V. Santiago and D. Guimarães. WEB-PerformCharts: A Collaborative Web-based tool for Test Case Generation from Statecharts. In *Proc. of iiWAS*, 2008, Linz, Austria, pages 374–381.
- [31] V. Santiago, M. F. Mattiello-Francisco, R. Costa, W. P. Silva, and A. M. Ambrosio. QSEE Project: An Experience in Outsourcing Software Development for Space Applications. In *SEKE '07*, pages 51–56, Skokie, IL, USA, 2007.