Editorial Board Member

Comments to the Author:

The reviewers appreciate the research conducted by the authors and believe the approach proposed to be promising. However, the submission requires a major revision before it can be accepted for publication. Specifically, the authors should improve on the following points:

\* formal definitions and notations (Reviewers 1 and 2)

\* usability of the approach: are the traces found the shortest ones possible (Reviewer 2), optimality of the solution (Reviewer 2) and its limitations (Reviewer 1)

\* improving the discussion: shortening/improving Section 6 (Reviewer 1), elaborating on "directed model checking" (Reviewer 2) and comparing the approach proposed with the existing solutions (Reviewer 2).

Referee(s)' Comments to Author:

# Reviewing: 1

Comments to the Author

SUMMARY

This paper presents an approach for bug reproduction that is based on model checking. Given a crash trace and a violated boolean property, the model checker is queried in a way so that it returns a counter example that corresponds to a path in the program from the entry to the crash location. The authors have developed a tool, JCHARMING, that chains WALA, Java Pathfinder, and a newly written generator for JUNIT tests. WALA helps reducing the program state space by reducing the program to a slice along the exceptional control flow, while the JUNIT generator provides the human tester with a reproducible bug.

The techniques themselves are not new, but their combination and application to the problem of bug reproduction is novel and quite promising. I recommend acceptance, but feel a few improvements are in order, namely

- proving a formal definition of the syntax of traces;

- proving a formal definition of a frame and a slice and a better discussion ("reasoning") of the algorithm;

- extending or modifying the evaluation section by a complete characterization

of the partially and not reproduced bugs in the 30-x cases (table)

- identifying, and making explicit, the limitations of JCHARMING

The review instructions also ask for possibilities for condensing (or amplifying) the text. In my opinion, the prose in section 6 could be shortened.

For more details, see below.

## Page 2

### l.16 "In [7]": poor style (writing rules say that the sentence must be complete if you leave out the reference), also see l.40

The sentence has been modified.

### l.21 "uses a list of function outputs": every output comes from a function -> be more specific. Also, make clear that JCHARMING additionally uses the source code.

Reviewer didn't get the meaning of the sentence here. output WHEN an uncaught...

### l.28 "control flow graph" seems a bit misleading since there is no 1-1 correspondence between statements in the source and program states.

??

### l.38 "Hadoop and, " -> Hadoop, and

The sentence has been modified.

### l.43 systems are "relatively complex": you contradict yourself a bit here, since, at least for complex bugs, you write in Section 7 that the quality of the trace is what matters.

The sentence has been modified. It now reads “*This is explained by the fact that the new systems used in this study are relatively larger compared to those [...]*”.

Yes, the quality of trace is important, but the size of the remaining slice is even more important.

# Page 3

### l.29 "NP-complex" does not exist as technical term IMO. A problem can be "NP-complete" or in the "complexity class@ NP.

NP-complex occurrences have been replaced by NP-complete.

### l.35-39: don't you mean that it is \_im\_possible to reproduce the bug?

Not impossible, one could create an approach that re-create a network packet

based on the crash characteristics.

### What does "specific data" refer to---simply the input?

All kinds of possible inputs (networks, disks access, GUI interactions, …)

# Page 4

### l.08 "minimal"... what?

The sentence sounds clear to me. Could you precise what you don't understand here ?

### l.39 typo in "JRapture"

The typo have been removed.

### l.42 "and on the replay": grammar

The grammar have been fixed.

## Page 5

### l.03 "code-level" delete hyphen

The hyphen has been deleted

### l.03 ", there exist": put in a new sentence

A new sentence has been created.

### l.28 "stack" -> stacks

The plural is on traces, i.e., stack traces.

### l.33 "prior to the work ... begins": grammar

The grammar has been fixed.

### l.34 add a period after "i.e."

A period has been added.

## Page 6

### l.42 "P is the set of properties that each state satisfies": incorrect or at least misleading scope of "the set." You mean "P associates with each state the set of properties that it satisfies."

The sentence has been reworked. It now reads : “*where S is the set of states, $T \subseteq S \* S$ represents the transitions between the states and $P$ associates with each state the set of properties that it satisfies.. The SUT is said to satisfy a set of properties p when there exists a sequence of states x leading towards these properties.*”

### l.47 what exactly is "x": a state or a \_sequence\_ of states (path)? I am familiar with the former, i.e., the situation where one asks the model checker whether property p holds in state x of model M, and then might get back the trace that constitutes a counter example. If "x" really is a path, then, for example, "x \in S" (Eq. 1) wouldn't hold, since $S$ is a set of states. Also, in Eq. (3), you would quantify over paths rather than states.

It's a state

### l.52 "specification" is not a good term here (and has a different meaning in formal methods), do you mean additional assumption?

The term specification has been removed. It now reads : “*In JCHARMING, SUTs are bound to a simple assumption:*”

### l.52 just as a remark, "fairness" has a special meaning in (LTL) model checking.

## Page 7

### l.17 "depicts" -> depict

The typo has been fixed.

## Page 8

### l.51 "are" -> is

The typo has been fixed.

## Page 9

### I don't understand Figure 5. Why is there a call to Foo.bar (even though in Figure 1 there is no edge from Bar.foo to Foo.bar)? What is "looptimes" - in Figure 1, the bound is called "loopCount"?

This is the result of the String constructor of InvalidActivityException in

java i.e. throw new InvalidActivityException("loopTimes");

### Where is the call to Bar.foo the "crash trace contains" (l.23).

In the "and 4 more" as explained in the Java documentation (http://docs.oracle.com/javase/6/docs/api/java/lang/Throwable.html) : *“This line indicates that the remainder of the stack trace for this exception matches the indicated number of frames from the bottom of the stack trace of the exception that was caused by this exception (the ``enclosing exception''). This shorthand can greatly reduce the length of the output in the common case where a wrapped exception is thrown from the same method as the ``causative exception'' is caught.”*

### Generally, I strongly recommend to include a grammar of a trace;

### this would also help answering some of my later questions (see below).

The grammar for java stacktrace is both verbose and very well-known by Java developers. This is why we didn't included it in the paper. However, we added a reference to an ANTLR (<http://www.antlr.org/>) V3 grammar for Java Stack traces written by Luca Dall'Olio, Christian Cassano and Gabriele Contini.

### l.38 typo in the caption; I would also reformat the caption to avoid the lonely "2."

The caption has been reformatted.

### l.42 what exactly is a frame? Put differently, what is the difference between a line number in a trace and a frame? Does f\_0 have to contain a property (in terms of a programming language: an expression), as Figure 5 and p. 12 suggest, or does it have to contain an exception?

As shown in Figure 5, we can see that the first line is the called frame 0 (referred to as frame f\_0), then subsequently the next line is called frame f\_1 , etc.

### l.46 "buggy ": delete the extra blank.

The extra blank has been removed.

## Page 10

### l.17 delete the extra "the."

The extra the has been removed

### l.21 "out of control" -> beyond control

The sentence has been modified.

### l.48 I am not sure whether I completely understand what a slice is in your context: does a slice consist of statements and declarations from the program source? And do frames constitute slicing criteria, or (since this wouldn't work) phrased more precisely: the method invocation they contain? Please provide a formal definition of a slice.

### l.54 typo in f\_{n 1}

The typo has been fixed

### l.55 in my understanding of a slice, the order of statements is essential, especially if the slice has to remain executable. Rather than forming the union of slides, you should therefore, e.g., concatenate them. (Perhaps the implementation is based on a union that happens to preserve the order, so that you never faced the difference between a set and a sequence). Algorithm 1 needs to be changed accordingly.

## Page 11

### l.09 "allow to reach": grammar

The grammar has been fixed.

### Figure 6: what is "entry"? What is the purpose of the two self-loops in z0 and z1?

Entry is the entry point of the program, i.e. the main in Java.

No purposes in particular, self-loops do happens

### l.31 as a comment, the "worst case scenario" is the best-case scenario from the perspective of state space explosion...

### l.44 how can a slice be empty? Since there is an underlying execution between frame f\_i and f\_{i+1}, there have to be program statements, no? Again, a definition of a slice would have helped me (perhaps). I also didn't understand why "offset" is reset to 1 in Algorithm 1.

### l.48 "would": is colloquial

### l.55 comma after WALA[31]

A comma has been added.

## Page 12

### l.6-l.14 what is the point in this paragraph? "At first sight, it may appear that static slicing alone be used..."... OK, and what at second sight?

The *second sight* is “*The result will be a dynamic slice. This assumes that we know in advance which input we need to provide to the program in order to reach the crash, which defeats the purpose of bug reproduction in the first place.”*

### l.20 "frame" -> frames

I couldn't find this typo.

### l.24 shouldn't "i+offset" be frames[i+offset]?Algorithm 1: shouldn't the slicing criterion change in each iteration of the loop?

### l.40, Eq.(6): I need some clarification on the premise of \models: assuming "x" is a path, what does "x." mean? Also, SUT and the union of slides are quite different types, why is it legal for both to write "x."?Further, what is the first subset relation good for (the one without the "x"-qualification)? Please introduce the notation and explain the different terms formally.

### l.45 What do you mean by "c\_{i>1}" needs to be included"? Is this an extra step or does the slicer find that "c\_{i>1}" automatically? And why can one include expressions anyway? Doesn't the slice contain statements?

### l.46 "needs to be untouched": I can't follow

If the first line of the stacktrace, also known as frame 0, has been tempered with, then, JCHARMING can't perform the slicing as it doesn't know where to go. The result is a non-directed model checking which is likely to fail.

## Page 13

### l.05 delete comma after "each"

The comma has been deleted

### l.10 delete "entry"

Entry has been deleted

### l.11 typo in s\_i+1

I couldn't find this typo.

### l.16 "As shown before": where?

A reference was missing. It now reads as shown in Figure 6.

# Page 14

### l.48 "an hypothetical" -> a hypothetical

The typo has been fixed.

### What are the limitations of the unlowering steps? Which parts can you not reconstruct? Please make explicit.

## Page 15

### Figure 8: faillure -> failure (3x)

Typos have been fixed.

### l.49 "Kilo Line": I've never seen KLoC spelled out like this.

It is spell KloC in many of our references. Which spelling would you prefer ?

### l.52 "make files": shouldn't it be Makefiles?

It now reads Makefiles.

## Page 16

### l.22 "player in" -> player among

I now reads player among.

### l.55 "built on the top" -> built on top

It now reads build on top.

### l.56 "that" -> , which

That has been replaced by which

## Page 19

### l.08, l.09: delete blanks before the setTarget... method (line 3 and 4 in trace, also in the trace on page 22)

The blanks have been removed.

## Page 20

### l.38: add comma after "class"

A comma was added.

### Removal of lines in the preprocessing step could also mean that variables relevant for control- or data dependence are not discovered, correct? This could then cause problem in the construction of test cases, correct?

Well, the removed lines belong to software that are outside the scope of the SUT such as the Java primitives or embedded libraries. If the bug lies there (outside the SUT) we can't reproduce it as JCHARMING won't find the point of entry to compute the static slice. So no, the preprocessing step cannot removed relevant control data or dependence.

## Page 21

### Both cases in Section 6.2 I don't find particularly enlightening. Surely, not match-able segments of traces present problems, but that is hardly surprising. Besides, in the second case, JCHARMING could employ a different definition of equality and in the first case include a heuristic. Instead of discussing what is more or less obvious, it would be much more interesting to count the frequency of those cases. Or to cluster resp. classify them, or to discuss how JCHARMING could improve its precision. Asked differently, what characterizes the "quality of ... crash trace" you refer to on page 23?

A qualitative stacktrace is a stacktrace where frames have not been tempered with (such as <deleted entry> in section 6.2 Partially Reproduced and where the frame 0 falls inside the SUT (such as 10052009 01:36:46 ERROR [Default: 1] struts.CPExceptionHandler.execute RID[(null;25KbxlK0voima4h00ZLBQFC;236Al8E60000045C3A 7D74272C4B4A61)] in section 6.3 Not reproduced.

### Several bug reports contain spelling mistakes (the two reports on page 21 contain 2 resp. 1 typo, the one on p.22f. another one). Are those in the original reports? (If so I'm not sure if one is allowed to correct them. lease inform yourself and get back to me :-)

Yes, these are the original reports. According to the APA guidelines : If “incorrect spelling, grammar, or punctuation in the source might confuse readers, insert the word sic, italicized and bracketed, immediately after the error in the quotation” (American Psychological Association, 2010, p. 172). For example, “they made they're *[sic]* lunches."

The typos and spelling mistakes in the report, however, doesn't confuse the readers, in our opinion.

## Page 22

### l.42-46 I can't follow your argumentation: why did JCHARMING not find the crash ocation? What is different in the case of Struts?

Because the SUT was Log4J and the error belongs to struts which is another software.

### l.47 capitalize log4j

Log4J has been capitalized.

### l.50 add comma after 46721

A comma has been added.

### l.50/51 "We believe that JCHARMING could have successfully..": why don't you just try it? Is it so complicated to rerun JCHARMING?

The generated exception do not belong to struts. Only the entry point is from struts. We'll have to do an undirected model checking here for the whole Log4J program. This undirected model checking will fail has for bugs 11570, 40212, 41186, 45335, 46271, 47912 and 47957 (Table II).

## Page 23

### l.27 add an "a" before "program's behavior"

A a has been added.

### l.50-54: correct, certain failures are "hardly reproducible." What to conclude, though?

## Page 24

### l.24 "bug" -> bugs

A s has been added.

### References (plenty of problems):

### Capitalization inconsistent in conference titles (e.g., [1,2,3,4]). Capitalization wrong in first names, tools, etc (e.g, [8,17,19]). Problems with non-ASCII letters (e.g., [15,25]). Incomplete reference (e.g., [14,17])

# Reviewing: 2

Comments to the Author

The authors present a technique to reproduce program crashes by using a combination of backward slicing and (directed) model checking.

The information obtained in the backward slicing operation serves as a way to guide the model checking activity. The approach is interesting,

and reasonably well explained, but the formal aspects need to be improved. Whenever formal notations are used, I got quite confused by the inconsistencies

that are currently in the text (see my detailed comments). I also would have liked to learn more about the way properties are expressed.

This is not explained at all in the current text.

Regarding the directly related work, I think the authors give a nice overview, but with respect to "directed model checking", the authors do not explain its origins at all (again, see detailed comments).

The main benefit of model checking over testing is its ability to determine that a system is absolutely correct. By starting with a bug and guiding the

checking to it, you are essentially not model checking anymore, but bug hunting. Of course, directed model checking in a way depends on there being a bug,

otherwise is devolves into traditional model checking. A discussion along those lines should be added, I think, to make this clear to the reader, since

"model checking" usually has this completeness characteristic of being able to show that a system is correct.

Can you say anything regarding the length of your counter-example traces? For understandability, it is crucial that counter-examples are short. It would

be interesting to know whether JCHARMING tends to produce the shortest counter-example or not, and if not, whether you could alter the technique to improve

this.

Can you produce results that show to what extent your approach is optimal? How many states were visited vs. how many absolutely need to be visited to find

the counter-example? Experiments like that would give an indication how well JCHARMING is able to find the counter-example based on the backward slice information.

Concluding, I think the article is promising and presents interesting results. However, I think the formal aspects need to be improved, and some more insight

could be given regarding the effectiveness of the approach.

Detailed comments:

## Page 1

### piece -> pieces

Piece is correct, THE most

## Page 2

### functions output -> function outputs

The typo have been corrected.

## Page 3

### oversee -> oversees

The typo have been corrected.

### a NP-Complex -> an NP-complete

It now reads NP-complete.

### different SMT (satisfiability modulo theories) solvers: different from what?

Different has been replaced by various. It now reads “*In order to overcome these limitations, some researchers have proposed to use various SMT (satisfiability modulo theories) solvers …”* .

### model checking techniques: SMT is being used for model checking itself,so the "even" remark seems to be odd.

The even has been removed. It now reads In order to overcome these limitations, some researchers have proposed to use various SMT (satisfiability modulo theories) solvers \cite{Dutertre2006} and model checking techniques \cite{Visser2003}.

### For example, the reading of a file that is only present on the hard drive of the customer or the reception of a faulty network packet: This is not a complete sentence. Please complete it or attach it to the previous sentence.

The sentence has been attached. It now reads : “*It is worth mentioning that both categories share a common limitation. It is possible for the required condition to reproduce a crash to be purely external, i.e., the reading of a file that is only present on the hard drive of the customer or the reception of a faulty network packet \cite{Chen2013a, Nayrolles2015}.”*

### reproduce a crash to be purely external.: so, is it possible to reproduce the crash / bug or not? You seem to indicate both.

The following clarifies this: “*Without this input, it may be challenging to reproduce the bug.”* The reproduction might still be possible but surely more challenging.

### for bug reproduction systems: also in [4], or in other papers?

## Page 4:

### file whose size averages 70KB -> files that are 70Kb on average

The typo has been fixed.

### Similarly, private: similar to the approach of Clause et al.? Please mention this

### clones: remove s

The typo has been fixed.

### JRaptrue -> JRapture

The typo has been fixed.

### creator -> creators

The typo has been fixed.

### it saw -> as

The typo has been fixed.

### phase: a proper ending of this sentence is missing, e.g. "are presented"

### that can also monitor other software system than the intended ones: why is that a problem?

It's violation of the host's privacy. You'll not only monitor your system but every Java systems running on the host.

### system -> systems

The typo has been fixed.

## Page 5

### level, there -> level. There

The typo has been fixed.

### lies: relies?

The typo has been fixed.

### really begins -> really beginning

The sentence sounds good to me

## Page 6

### used based -> based

The typo has been fixed.

### Except for STAR: so how does JCHARMING compare to STAR?

We couldn't compare to STAR. STAR's source code is available but do not compile. We will have to re-implement the whole approach to compare both tools. Accounting for programming mistakes and imprecision about the approach in Chen's Ph.D thesis, the comparison wouldn't be accurate.

### the system: for model checking, it is crucial to mention that the system is formally defined (has a clear semantics)

The sentence now reads : “*Model checking (also known as property checking) will, given a formally defined system (that could be software \cite{Visser2003} or hardware based \cite{kropf1999introduction}), check if the system meets a specification by testing exhaustively all the states of the system under test (SUT), which can be represented by a Kripke \cite{Kripke1963} structure:*”

### - P is the set of properties that each state satisfies → P is a set of state predicates, and there should be a function assigning subsets of P to states. Typically, however, P does not hold in its entirety in all reachable states

The sentence has been reworked. It now reads “*where S is the set of states, $T \subseteq S \* S$ represents the transitions between the states and $P$ is a set of state predicates, and there should be a function assigning subsets of P to states (P does not necessarily hold in all reachable states).*”

### The SUT is said to satisfy a set of properties: you seem to give a definition here of satisfying a property that corresponds with testing. For model checking, a property is said to hold if there exists no trace contradicting it.

Indeed. The sentence has been reworked and now reads : “*The SUT is said to satisfy a set of properties $p$ when there exists no sequence of states $x$ leading to a state contradicting $p$”*.

### not that p holds nor that ∀x, p is satisfiable: if that is the case, then the system does not satisfy p.

Yes, p refers to a state where the program crashes. Therefore, we don't want the program to satisfy p.

### fair environment: is there any relation to fairness in formal verification, as in "Behaviour that is infinitely often enabled is also infinitely often executed"?

### properties: How are the properties expressed, using which temporal logics?

In Java Pathfinder (the model checking engine we extended), the system is exercised until it crashes. If no crash can be yield, then the property *The program must not crash* is satisfied. There is only one property that we don't need to define. The plural here is only use in the context of preliminaries about model checking.

## Page 7:

### - (3): this is closer to satisfying a property in a model checking context. Usually, the "for all x" does not need to be added though. In fact, the shorter notation SUT |= p would be more in line with what is common

The equation has been modified accordingly to your suggestion.

### such a path exists -> there exists a path contradicting this

The sentence has been reworked and now reads: “*If there exists path contradicting $p$ (i.e., $\exists x$ such that $(SUT, x) \models c$) then the model checker engine will output the path $x$ (known as the counter-example), which can then be executed. The resulting Java exception crash trace is compared with the original crash trace to assess if the bug is reproduced.*”

### depicts -> depict

The typo has been fixed.

## Page 8:

### what to look for in order to detect the causes: do you mean intermediately, i.e. along the trace? Because the final goal is known here (i > 2)

### complete, but impractical: that depends on the search order, which, as you mention later, opens the door for directed model checking

### directed (or guided) model checking has been introduced [28]: actually, the term "directed model checking" was coined in an earlier paper, namely "Directed Explicit-State Model Checking in the Validation of Communication Protocols", by Edelkamp, Leue, and Lluch-Lafuente. Also, you should refer to "Survey on Directed Model Checking" by Edelkamp, Schuppan, Bosnacki, Wijs, Fehnker and Aljazzar

Both references have been added.

### use insights -> uses insights

The typo has been fixed.

## Page 9:

### Figure 5: Bar.Goo -> Bar.Foo. Also, it may improve readability if you add f0, f1 etc to the figure, as you refer to these in the text

The typo has been fixed.

## Page 10:

### that led to the: remove the

The sentence sounds right to me : “*In such a case, we want to reproduce the root exception (line 28) that led to the other two (lines 14 and 1). This said, we remove the lines 1 to 14.”*

### are not limited to, -> but not necessarily, to

The typo has been fixed.

### also reduce -> reduce

The typo has been fixed.

### foo: previously, "foo" was capitalised. Please be consistent

The typo has been fixed.

### foo: see previous statement

The typo has been fixed.

## Page 11:

### - equation 5: I have a number of comments regarding (5):

### \* i ranges from 0 to entry, but f\_entry was previously not defined, only just entry

I couldn't find any reference to f\_entry in equation 5.

### \* i should range to entry-1, since otherwise there should also be an f\_entry+1, and this appears not to be the case

This has been corrected.

### \* equation 4 states that the union equals the bslice from f\_0 to entry, here it is a subset. Which one is correct?

Both. A backward slice (bslice) is a subset.

### Indeed, in Figure 6, the set of states: based on figure 6, this set should be empty, as f\_2 has no outgoing transitions. The other sets you mention (f2 to f1 and f1 to f0) are also empty.

The states were inversed. The sentence now reads “I*ndeed, in Figure \ref{fig:jcharming-slice}, the set of states allowing to reach $f\_2$ from $f\_0$ is greater than the set of states to reach $f\_2$ from $f\_1$ plus set of states to reach $f\_1$ from $f\_0$.”*

### assuming that z2 is a prerequisite: what is a prerequisite?

I prerequisite is a state where you must pass.

### Which state in figure 6 corresponds with entry?

F2.

### Why is bslice from f\_0 to entry the given set of states?And the sentence ends prematurely. Which set corresponds with the union of the given bslices? This example is very confusing. Please fix this.

### between each frame -> between each two frames

This has been corrected.

### From line 1 to line 5: in Alg. 1, please number the lines, as you use these numbers in the text. Also the lines in the text are wrong. Instead of "line 1 to 5", you should write "line 1 to 6", and instead of "line 6 and ends at line 15" you should write "line 7 and ends at line 15".

The lines have been numbered and the line number fixed in the algorithm's description.

### then JCHARMING: remove then

The then has been removed

### the possibility to resort to non-directed model checking: and what if the final slice is not empty? How does it help the directed model checking? This is explained later, but it would improve the text if you already give some indication how this works here.

### provides -> provide

The typo has been fixed.

## Page 12:

### Frames frames ← extract frames from crash stack;: do not write text in math mode

The text is no longer in math mode.

### size of frame: which frame? Multiple were extracted in previous line. Or do you mean frames?

Yes, I meant frames. This typo has been corrected.

### equation 6: I have the same issues with this equation as with equation 5. In addition:

### \* are the two cases between the brackets conjunctives? How are they combined?

### \* what is the difference between the two cases? How should I read the second one?

A comma was missing in the equation. I believe it's easier to understand now.

### both the transitions and the states are entry elements: how can this be interpreted from the equation?

See previous answer.

### That is the -> That is, the

A comma has been inserted here.

### only frame that needs to be untouched for the backward static slice to be meaningful is f0.: can you elaborate on this? Why is this the case?

F0 is the crash point. If the frame containing the crash point is corrupted or missing, then we can't perform the slicing (as we don't have any direction to point to). If this is the case, the slicing will fail and we will fall back into undirected model checking.

## Page 13:

### choose -> chose

The typo has been fixed.

### each, forward -> each forward

The typo has been fixed.

### with entry the states: something is wrong here. Please fix

I couldn't find this.

### that falls -> that fall

The typo has been fixed.

### a set a property -> either "a property" or "a set of properties"

The typo has been fixed.

### transitions -> transition

The typo has been fixed.

### t is the percentage: do not start a sentence with a mathematical symbol

The sentence now reads : “Thus, $t$ is the percentage of identical frames between both crash traces.”

### the exercise of the bug -> execution of the bug, or "the bug to appear"

The sentence now reads “*To help software developers reproduce the crash in a lab environment we automatically produce the JUnit test cases necessary to run the SUT to cause the bug to appear.*” as suggested.

## Page 14

### industrial size -> industrial sized

The typo has been fixed.

### listens the -> listens to the

The typo has been fixed.

## Page 15:

### a loop from 0 to 3: what do you mean by this?

The sentence now reads “*Then, the jsme.Bar.foo(int) method will execute a $for-loop$ from $i=0$ until $i=3$ and [...]”*

### as shown in Figure 8.: please discuss figure 8

Figure 8 as been described. The paragraph now reads : “The generation of the JUnit test itself is based on templates and targets directly the system under test. Templates are excerpts of Java source code with well-defined tags that will be replaced by values. We use templates because the JUnit test cases have common structures. Figure 8 shows our template for generating JUnit test cases. In this figure, each { % % } will be dynamically replaced by the corresponding value when the JUnit test is generated. For example, { % SUT % } will be replaced by Ant if the SUT is Ant. First, we declare four variables that contain the faillure, the threshold above which a given bug is said to be partially reproduced, the differences which count how many lines differ between the original failure and the failure produced by JCHARMING and a StringTokenizer. The StringTokenizer allows to break the original failure into tokens. Second, the test method where { %SUT% } is replaced by the name of the SUT and { %STEPS% } by the steps to make the SUT crash. Then, the stack trace related to the crash is received in catch part of the try-catch block. In the catch part, we compute the number of lines that do not match the original exception and store it into differences † .Finally, the assertTrue call will assert that the stack traces from the induced and the original crash are at least threshold percent similar to each other.”

### study -> studies

The typo has been fixed.

### reasonable amount of time: what is a reasonable amount of time?

A reasonable amount of time refer to a task that can be done several time a day by a developer, using its own hardware, without jeopardizing its workday. The point of JCHARMING is to speed-up the reproduction process and therefore the creation of a fix. Consequently, we don't want our process to take hours to complete.

## Page 16:

### time JfreeChart -> time. JfreeChart

The typo has been fixed. The sentence now reads : “*JfreeChart [39] is a well-known library that enables the creation of professional charts. Similar to dnsjava, it has been maintained over a very long period of time—JfreeChart was created in 2005— and it is a relatively large application.*”

### on the top -> on top

I couldn't find this.

### Mahout: in the table, swap the entries for Hadoop and Mahout, to keep it in line with the order in the text

Mahout and Hadoop have been swapped in Table I.

## Page 17:

### above t=80%: and below 100%, I assume

The sentence now reads : The result is ``Partial'' if the similarity between the crash

trace generated by the model checker and the original

crash trace is above t=80\% and below t=100\%.

### fill out all -> fill all

The typo has been fixed.

## Page 18:

### dispatching:”: end sentence with '.'

This a quote from the original bug report where the sentence ends with a “:”. Furthermore, the “:” introduces the stack traces to come.

## Page 19:

### neither -> either

The typo has been fixed.

### nor -> or

The typo has been fixed.

## Page 20:

### hour -> hours

The typo has been fixed.

## Page 21:

### ChecksumException Re-running -> ChecksumException. Re-running

A point has been added.

### the attached -> of the attached

The typo has been fixed.

### While, -> While

The typo has been fixed.

### were -> was

The typo has been fixed.

## Page 22:

### of few -> of a few

The typo has been fixed.

### Application -> Applications

The typo has been fixed.

### We believe that JCHARMING could have successfully reproduced the crash: couldn't you test this?

## Page 23:

### the SUTs analyzed by JCHARMING are the same as the ones used in similar studies: how do the results of JCHARMING actually compare to those obtained in similar studies?

### footnote "described here ... ": make this note a full sentence

The footnote now only contains the link.

## Page 24:

### all based on -> all written in

The typo has been fixed.

### bug -> bugs

The typo has been fixed.

### involves -> involve

The typo has been fixed.

### Katoen Jp -> Katoen, J-P.

## Page 26:

### 41. Hadoop A. Hadoop 2011.

### 42. Mahout A. Scalable machine learning and data mining 2012.

### 43. Snyder B, Bosanac D, Davies R. ActiveMQ in Action 2011; :408.: these are not very helpful references. Please add more information