



UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

Realtime data collection in IDEs

by

David Chenaux

Thesis for the Master of Science in Computer Science
Supervised by Prof. Dr. Philippe Cudré-Mauroux

eXascale Infolab
Department of Informatics - Faculty of Science - University of Fribourg

January 9, 2017

UNIVERSITY OF FRIBOURG

Faculty of Science

Department of Informatics

eXascale Infolab

Thesis for the Master of Science in Computer Science

Supervised by Prof. Dr. Philippe Cudré-Mauroux

by David Chenaux

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Contents

Abstract	2
Acknowledgements	3
List of Figures	6
List of Tables	7
1 Introduction	8
1.1 Problem definition	8
1.2 Goals and objectives	9
1.3 Organization	9
2 Related work	10
2.1 What is Program Analysis ?	10
2.2 Program Analysis approaches	12
2.2.1 Static methods	12
2.2.2 Dynamic methods	12
2.3 Program Analysis tools	12
2.3.1 SPA tools	13
2.3.2 DPA tools	13
2.4 Current restrictions	13
3 Development	14
3.1 Proposed solution	14
3.2 Environement	14
3.3 Data capture model	14
3.4 Data model	15
3.5 User interface	15
4 Installation guide	16
4.1 Setup the environment	16
4.2 Use the packaged version	16
4.3 From source code	16

5	Experiments	17
5.1	Test script	17
5.2	Data extraction times	17
5.3	Writing times to database	17
5.4	Reading times from database	18
6	Conclusion	19
6.1	Conclusion	19
6.2	Future work	19
A	Glossary	20
B	License of the software	21
	Bibliography	22

List of Figures

List of Tables

2.1	Comparison of Dynamic analysis with Static Analysis	11
-----	---	----

Chapter 1

Introduction

Integrated development environments (IDEs) have been around for a few decades already, yet none of the modern IDEs was able to successfully integrate their source code editors with the actual data stream flowing through the code. Ability to display the actual data running through the system promises many potential benefits, including easier debugging and code recall, which results in significantly lower code maintenance costs.

1.1 Problem definition

Every developer is more or less feared about the debugging and code reviewing phase of their software. Obviously, this process can sometimes take several painfully hours and each programmer knows how frustrating it can be to search for a hidden bug in thousands lines of codes. In order to support the programmers in this hated task, debuggers are the most useful existing tools which are part of the so called *static program analysis* (SPA).

With the apparition of object-oriented programming language, searching for syntactic errors in the code is not anymore sufficient. Therefore, a new research field was pushed forward which is called the *dynamic program analysis* (DPA) and consists in analyzing the software during it's execution. This procedure allows to take in account some possible inputs which weren't probed with the SPA. Yet none of the modern IDEs was able to successfully integrate their source code editors with the actual data stream flowing through the code. This is why the present project, which goals and objects are defined in the next section, is aiming to contribute to the subject.

1.2 Goals and objectives

The goal of this project is to design a proof-of-concept system in one programming language that allows full code instrumentation. This system should be able to seamlessly capture all values for all variables in source code and store them somewhere, with further possibility to easily retrieve saved values. The system should also provide an API to the storage in order to make the data accessible for navigation and display in third-party applications. Also, a basic visualizing interface will also be included in order to allow an easy review of the results. Finally, an evaluation of system's performances will be established through different experiments.

1.3 Organization

The thesis is divided in four main sections :

1. **Related work:** In this first chapter of the thesis, an insight of the existing work on the field *program analysis* will be presented and in particular the [DPA](#). This is including a definition of the field and its particularities, an overview of some available solutions side by side with the current restrictions.
2. **Development:** This part is focusing on the development of the proof-to-concept system with a presentation of the proposed solution and detailed information about its structure.
3. **Installation guide:** Simply an installation guide of the software which describes the needed environment, the package installation and the compilation of the system.
4. **Experiments:** Finally in this section a few experiments will be conducted in order to test and check the performance and results of the software.

The thesis concludes with some outputs and is proposing some future improvements which seem to be important.

Chapter 2

Related work

“Sharing is good, and with digital technology, sharing is easy.”

Richard Stallman

As stated in the previous chapter, the goal of this thesis is to implement a dynamical program analysis system. In order to build a theoretical background, program analysis will be defined in this chapter along with the presentation of some available solutions and their restrictions.

2.1 What is Program Analysis ?

Programming environments are an essential key for the acceptance and success of a programming language. After [Ducassé and Noyé \[1994\]](#), without the appropriate developments and maintenance tools, programmers are likely to have a bad software understanding and therefore produce low-quality code. They will be therefore reluctant to use a language without appropriate programming environments, however powerful the programming language is.

As already introduced in the previous chapter, program analysis is an automated process which aims to analyze the behavior of a software regarding a property such as correctness, robustness, safety and liveness. Program analysis can be separated in two methods : the [SPA](#) which is performed without running the software and the [DPA](#) which is obviously fulfilled during runtime. [\[Wikipedia, 2016\]](#)

The [SPA](#) is a really simple solution because it does not require running the program for analyzing the dynamic behavior of a program. It consists in going through the source code and highlight coding errors or ensure conformance to coding guidelines. A classic example

of static analysis would be a compiler which is capable of finding lexical, syntactic and even semantic mistakes. The main advantage of this method is that it allows to reason about all possible executions of a program and gives assurance about any execution, prior to deployment.

Nevertheless, according to [Gosain and Sharma \[2015\]](#), since the widespread use of object oriented languages, SPA is found to be ineffective. This can be explained because of the usage of run-time features like dynamic binding, polymorphism, threads etc. To remedy this situation, developers call on DPA which can, after [Marek et al. \[2015\]](#), gain insight into the dynamics and runtime behavior of those systems during execution. Because the runtime behavior depends on many other factors, such as program inputs, concurrency, scheduling decisions, and availability of resources, static analysis is not capable of retrieving those values. The following table, proposed by [Gosain and Sharma \[2015\]](#), is resuming the main differences between static and dynamic analysis.

Dynamic Analysis	Static Analysis
Requires program to be executed	Does not require program to be executed
More precise	Less precise
Holds for a particular execution	Holds for all the executions
Best suited to handle run-time programming language features like polymorphism, dynamic binding, threads etc.	Lacks in handling run-time programming language features.
Incurs large run-time overheads	Incurs less overheads

TABLE 2.1: Comparison of Dynamic analysis with Static Analysis

A relevant point which comes out of this comparison, is that Dynamic Program Analysis is not replacing the Static Analysis, but on the contrary it is a complementary tool. Indeed, even if Static Program Analysis is not sufficient anymore, it gives nevertheless important information about the code for the programmer. The DPA is coming in a second phase when the SPA has been processed and the errors corrected. As it can be deduced, the main advantage of DPA is that it can examine the actual and exact run-time behavior of the program, whereas SPA main advantage is that it does not depend on input stimuli and can be generalized for all executions. To illustrate these differences, some program analysis solution will be presented further in this chapter.

2.2 Program Analysis approaches

Now that a definition of Program Analysis has been established, some different approaches have to be exposed in order to fully understand the subject. Nevertheless as the field is really vast, it is not the aim to cover the entire subject. Yet the reading of this section should give a good overview to the reader. First, the main static analysis methods will be steered following logically with the dynamic analysis methods.

2.2.1 Static methods

The static methods are regrouped in four different categories proposed by [Nielson et al. \[2004\]](#) and presented here.

Data Flow Analysis: is a technique which consist in gathering information about the values and their evolution at each point of the program. In the Data Flow Analysis the program is considered as a graph in which the nodes are the elementary blocks and the edges describe how control might pass from one elementary block to another.

Constrained Based Analysis: or Control Flow Analysis, is a method which aims to know which functions can be called at various points during the execution of a program.

Abstract Interpretation:

Type and Effect Systems:

2.2.2 Dynamic methods

Testing:

Monitoring:

Program slicing:

2.3 Program Analysis tools

This section is dedicated to the available solutions in terms of program analysis. As it will be explained in the next chapter, the proof-to-concept system will be coded in *Python* and therefore a special focus will be made on solutions available in this language. First,

some Static Analysis solutions will be presented following with the dynamic method ones.

2.3.1 SPA tools

One of the more complete and popular Static Program Analysis tool in python is PyLint.

Instrumentation based VM Profiling based Aspect Oriented Programming

2.3.2 DPA tools

Pythontutor

2.4 Current restrictions

Chapter 3

Development

“For me, open source is a moral thing.”

Matt Mullenweg

In this chapter, the proposed solution will be explained in details including the Setup, Data capture model, Data model and it's User interface.

3.1 Proposed solution

start with the detailed description of what we actually propose

3.2 Environement

The project is articulated around 4 main technologies. First the data is captured in *Python* with the help of the integrated Debugger Framework. Then the extrated data is stored in a *MongoDB* Database. Finally they are processed and showed with the help of *Python*, *Html/CSS* and *Javascript*. Each module of the solution is presentend in details in the following sections.

3.3 Data capture model

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in

reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.4 Data model

how you store data in a database

3.5 User interface

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 4

Installation guide

“If Microsoft ever does applications for Linux it means I’ve won.”

Linus Torvalds

In this chapter, the complete installation process of the developed script will be presented.

4.1 Setup the environment

First format windows and install linux because you are not a loser

4.2 Use the packaged version

In order to simplify the installation process, a packaged version has been built and is ready to download on the project’s [GitHub page](#).

The installation process is really straightforward and since it’s a [pip](#) package.

4.3 From source code

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 5

Experiments

In this section different type of experients will be conducted in order to test and check the performance of the developped software. In order to test the following variables the same script was used for all experiments.

5.1 Test script

In order to conduct the different experiments, a test script has been chosen.

5.2 Data extraction times

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5.3 Writing times to database

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur

sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5.4 Reading times from database

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 6

Conclusion

6.1 Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

6.2 Future work

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Appendix A

Glossary

DPA Dynamic Program Analysis

pip Pip Installs Packages is a package management system used to install and manage software packages written in Python

SPA Static Program Analysis

Appendix B

License of the software

Copyright (c) 2016 DAVID CHENAUX

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Bibliography

Mireille Ducassé and Jacques Noyé. Logic programming environments: Dynamic program analysis and debugging. *The Journal of Logic Programming*, 19:351–384, 1994.

Anjana Gosain and Ganga Sharma. *A Survey of Dynamic Program Analysis Techniques and Tools*, pages 113–122. Springer International Publishing, Cham, 2015. ISBN 978-3-319-11933-5. doi: 10.1007/978-3-319-11933-5_13. URL http://dx.doi.org/10.1007/978-3-319-11933-5_13.

Lukáš Marek, Yudi Zheng, Danilo Ansaloni, Lubomír Bulej, Aibek Sarimbekov, Walter Binder, and Petr Tůma. Introduction to dynamic program analysis with disl. *Science of Computer Programming*, 98, Part 1:100 – 115, 2015. ISSN 0167-6423. doi: <http://dx.doi.org/10.1016/j.scico.2014.01.003>. URL <http://www.sciencedirect.com/science/article/pii/S0167642314000070>. Fifth issue of Experimental Software and Toolkits (EST): A special issue on Academics Modelling with Eclipse (ACME2012).

F. Nielson, H.R. Nielson, and C. Hankin. *Principles of Program Analysis*. Springer Berlin Heidelberg, 2004. ISBN 9783540654100.

Wikipedia. Program analysis — wikipedia, the free encyclopedia, 2016. URL https://en.wikipedia.org/w/index.php?title=Program_analysis&oldid=732080552. [Online; accessed 7-January-2017].