SIMONE: Architecture-Sensitive Near-miss Clone Detection for Simulink Models

James R. Cordy

School of Computing, Queen's University
Kingston, ON, Canada K7L 2L8

cordy@queensu.ca

ABSTRACT

Identification of similar fragments of software systems, or clones, has many applications in software engineering and maintenance, including quality control and improvement, standards compliance, test management and failure analysis and prevention. Code similarity analysis systems, or clone detectors, are a mature and widely used technology in traditional software code maintenance. As model driven engineering continues to advance, technologies such as Simulink are increasingly widely used to design and implement automotive software systems. Automotive Simulink models are particularly prone to cloning due to the copy-paste authoring paradigm of the Simulink IDE, and the inherent similarity of elements and tasks in automotive applications. Thus the ability to find Simulink model clones is equally important, but is much less thoroughly studied and used.

Simulink models can be viewed as graphs, and subgraph isomorphism is the most obvious technique to implement model similarity analysis. While graph-based model clone detection techniques show good results in finding exactly similar subgraphs in graphical models, they have difficulty in finding near-miss matches, which may vary by incidentally added or removed blocks, lines, inputs or outputs, and largely ignore the architectural structure of the model, making analysis results difficult to communicate to practitioners.

In this talk I will introduce SIMONE, a Simulink model clone detector designed to address these issues. SIMONE is a hybrid architecture / text-based clone detector that yields actionable results that can be adopted directly into the modelling engineer's workflow as part of their everyday Simulink IDE interactions. SIMONE uses a two-stage analysis, beginning with the extraction and normalization of the architectural elements (subsystems) to be analyzed, and then uses the mature text-based near-miss code clone detector NICAD to efficiently identify structurally meaningful near-miss subsystem clones. By transforming the graph-based models to normalized text form for comparison, SIMONE is able to uncover important model similarities that are difficult to find in any other way. SIMONE is domain-specific to Simulink / Stateflow models, and directly integrates its results into the modelling engineer's own working environment in the Simulink IDE.

I will begin by outlining the problem of Simulink model clone detection, and walk through the challenges we faced in designing and implementing SIMONE. Following a quick comparison with other model clone detectors, I will demonstrate the application of SIMONE to both the high level and low level analysis of industrial automotive models.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WASA'15, May 4, 2015, Montréal, QC, Canada.
Copyright © 2015 ACM 978-1-4503-3444-0/15/05 ...$15.00.
http://dx.doi.org/10.1145/2752489.2752498.

Categories and Subject Descriptors

D.2.2 [**Software Engineering**]: Design Tools and Techniques—*model driven engineering*; D.2.7 [**Software Engineering**]: Distribution, Maintenance, and Enhancement—*reverse- and re-engineering*

General Terms

Design, Measurement, Verification

Keywords

Model-driven engineering, Software clones, Simulink

ABOUT THE SPEAKER

James Cordy is Professor and past Director of the School of Computing at Queen's University in Kingston, Ontario, Canada. From 1995 to 2000 he was Vice President and Chief Research Scientist at Legasys Corporation, a software technology company specializing in legacy software system analysis and renovation.

Dr. Cordy received his PhD from the University of Toronto, where he served several years as lecturer and senior research associate before moving to Queen's University in 1985. As leader of the TXL source transformation project with hundreds of academic and industrial users worldwide, he has been involved in software analysis and transformation systems for more than 20 years.

He has published more than 170 refereed contributions in software engineering, programming languages and artificial intelligence. Most recently he leads the Model Pattern Engineering project of the NECSIS Automotive Partnership, an academic-industrial collaboration involving General Motors, IBM Canada, Malina Software Corp. and several Canadian universities.

Dr. Cordy serves widely as member and chair of conferences and workshops in programming languages and software engineering, recently chairing the IEEE 2011 International Conference on Software Maintenance, the IEEE 2012 International Working Conference on Source Code Analysis and Manipulation, the 2012 International Working Conference on Reverse Engineering, the 2012 Dagstuhl Workshop on Software Clone Management in Industrial Application, and the IBM Centre for Advanced Studies 23rd International Conference on Computer Science and Software Engineering. He is an ACM Distinguished Scientist, a Senior Member of the IEEE, and an IBM Centre for Advanced Studies Visiting Scientist and Faculty Fellow.

ACKNOWLEDGEMENTS

This work is supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) as part of the NECSIS Automotive Partnership with General Motors, IBM Canada and Malina Software Corp., and by an Ontario Research Fund Research Excellence grant.

1. REFERENCES

[1] M. H. Alalfi, J. R. Cordy, T. R. Dean, M. Stephan, and A. Stevenson. Models are code too: Near-miss clone detection for Simulink models. In *28th IEEE Int. Conf. on Softw. Maint., ICSM 2012*, pages 295–304, 2012.

[2] J. R. Cordy. Submodel pattern extraction for Simulink models. In *17th Int. Software Product Line Conf., SPLC 2013*, pages 7–10, 2013.

[3] J. R. Cordy and C. K. Roy. The NICAD clone detector. In *19th IEEE Int. Conf. on Program Comprehension, ICPC 2011*, pages 219–220, 2011.

[4] C. K. Roy and J. R. Cordy. NICAD: Accurate detection of near-miss intentional clones using flexible pretty-printing and code normalization. In *16th IEEE Int. Conf. on Program Comprehension, ICPC 2008*, pages 172–181, 2008.