Bavaria California Technology Center (BaCaTeC)
Internationalization of the High-Tech-Initiative

Proposed Project (Two pages maximum – including this page)

Short Title of Project: Improving symbolic software verification via machine learning

Category (tick off)**:**

o life sciences X i&c technologies o new materials o environmental technologies o mechatronics

Contact Person BY:
name, title: Prof. Dr. Gerald Lüttgen
institution: Universität Bamberg
(subinst.:) Lehrstuhl Softwaretechnik und Programmiersprachen
full address: 96045 Bamberg
phone: +49-951-863-3850
fax: +49-951-863-3855
e-mail: gerald.luettgen@swt-bamberg.de
web: http://www.swt-bamberg.de/luettgen/

Contact Person CA:
name, title: Prof. Gianfranco Ciardo
institution: Univ. of CA, Riverside
(subinst.:) Comp. Sci. & Eng. Dept.
full address: 351 Winston Chung Hall
University of California
Riverside, CA 92521
phone: +1-951-827-2281
fax: +1-951-827-4643
e-mail: ciardo@cs.ucr.edu
web: http://www.cs.ucr.edu/~ciardo/

Preparation of Cooperation: Prof. Lüttgen chairs the Software Technologies Group at Otto-Friedrich-University Bamberg since its founding in 2009. Bamberg's Department of Information Studies and Applied Informatics is expanding into Computer Science, where Prof. Lüttgen's group plays a key role with its research in Concurrency Theory and Automated Verification.
Prof. Ciardo is a Full Professor at UC Riverside since 2003 where he heads a research group in the modeling and analysis of stochastic systems, focusing on Petri Nets and Model Checking. UC Riverside is one of ten campuses that comprise the University of California system which is widely recognized as the preeminent public university system in the world.
Profs. Lüttgen and Ciardo have a strong history of collaboration between 2000 and 2007, which started when Prof. Ciardo was on the faculty at William and Mary (Williamsburg, VA, USA) and Prof. Lüttgen was a Research Scientist at ICASE (NASA Langley, Hampton, VA, USA). Their collaboration has resulted in several conference and journal publications on aspects of automated verification: *symbolic state-space generation* (ICATPN 2000, TACAS 2001, CAV 2007, FMSD 2007, STTT 2009), *model checking* (TACAS 2007), and *static variable ordering* (ICATPN 2007).
This proposal is to jump-start a new collaboration between the research groups of Prof. Lüttgen and Prof. Ciardo, and has been initiated by a visit of Prof. Ciardo to Bamberg in September 2012.

Summary of Proposed Research/Cooperation (*)**:** Given society's increasing dependence on digital systems, the ability to assess their correctness is essential to achieve marketplace success, reduce deployment risks and guarantee safety and security. For hardware, so-called *symbolic model checking* based on decision diagrams (DDs) has established itself as a successful approach to verifying complex digital circuits. For software, however, much further research is needed since state spaces describing program behavior are typically much larger than those of circuits; thus, state-space exploration can overwhelm even the best currently-known algorithms. One weakness of DD-based algorithms is that DD sizes are much affected by the order in which system variables are considered, and determining an optimal order is known to be extremely difficult (NP-hard). Another related but rarely explored problem is that the order in which system events are applied

(*)Applications with explicitly exposed contribution of the Bavarian and the Californian Partner are preferred.

(**) The total funding – if approved – will be available for **18 months** maximum. It starts **July 1st** for April applications and **January 1st** for October applications respectively. It will be split into two years and extra defined for each calendar year. So the application must be displayed for each calendar year separately. Unused financial means are retracted at the end of each calendar year.

when exploring state spaces also affects the sizes of the DDs during exploration. Our *saturation* heuristic to guide state-space exploration (TACAS 2001) considers this aspect, saving many orders of magnitude in memory and time. However, to target software systems of practical interest, we need a radical new heuristics-based approach that not only greatly increases the size of systems that can be verified, but also lets engineers use model checking tools without having to decide critical parameter settings on variable and event orderings.
In this project we will (a) systematically study the influence of such critical parameters on numerous benchmark models taken from the literature and industrial case studies, and (b) explore how machine learning techniques can help determine 'good' parameter settings for a given verification problem by analysing partial runs on smaller instances of the same problem. This has become feasible due to the availability of powerful *multicore systems* which will let us execute many symbolic model checking problem instances in parallel and gather data such as DD sizes and shapes, event locality (i.e., how different events can affect different portions of the DD) and runtimes. We will then feed this data to *machine learning* techniques to correlate problem instance features with successful variable and event orders, thus providing a basis for novel heuristics for symbolic model checking. This should greatly outperform previous approaches that simply consider the DD without taking into account the nature of the model under study.

Expected Output: The immediate project results, which shall be submitted to leading international conferences and journals, are twofold. Firstly, we will gain a deeper understanding of why and when symbolic algorithms work best. Secondly, we will provide the groundings for a framework of high-performance model-checking algorithms that can learn from partial runs and optimize their own parameters to minimize computing resources.
The most important outcome, however, will be a long-term collaboration between both research groups that shall be sustained via joint DFG/NSF multi-year grants. The first grant proposal will build upon the research results from this project and shall support multiple PhD students/Postdocs.

Travelling Scientists (Junior/Senior): *From Bamberg:* Prof. Dr. Gerald Lüttgen (senior); Dr. David White (senior, a Postdoc expert in machine learning who received a PhD in this field from the University of York, UK); Master student (to be named, assisting the Bamberg group in the context of his/her Master project). *From UCR:* Prof. Gianfranco Ciardo (senior); Ms. Xiaoqing Jin (PhD student, who is conducting research on parallel DD-based model checking techniques).
Involvement of graduate students: 40 % (between 0 and 100% – for BaCaTeC-statistic)

Financial contribution of involved company: Not applicable since no company is involved.

Financial Requests for travel expenses (**)**:**
1st calendar year (2013) total: **6000 €**
2nd calendar year (2014) total: **4000 €**

Detailed assembly of expenses: In 2013, 6000 € are estimated to cover accommodation costs for one 1 week visit of Dr. White (500 €) to Riverside, travel and accomm. expenses for one 2 week visit of the Bamberg Master student (1500 € + €1000) to Riverside, accomm. expenses for one 1 week visit of Prof. Ciardo (500 €), and travel and accomm. expenses for one 2 week visit of Ms. Xiaoqing (1500 € + 1000 €) to Bamberg. In 2014, one further 1 week visit of both Dr. White to Riverside (1500 € + 500 €) and Prof. Ciardo to Bamberg (1500 € + 500 €) are planned, for an estimated total of 4000 €. The travel of Dr. White in 2013 and the expenses of 1 or 2 short visits of Prof. Lüttgen to Riverside in 2013/2014 will be paid out of the general budget of Prof. Lüttgen's research group. Prof. Ciardo's travel costs in 2013 will be covered by UC Riverside.