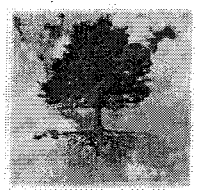

European Commission
Information Society and Media

Scientific Review Report
Period from 01/02/2010 to 31/01/2011

Project Acronym: CERCO
Grant agreement number: 243881
Project starting date: 01/02/2010
Project duration: 36 months

Project web site: http://cerco.cs.unibo.it/
Coordinator: Dr Claudio Sacerdoti Coen
Date and place of review meeting: 11 March 2011, Bologna

Name (s) of expert(s) and scientific officer:

Prof. Sandrine Blazy
Dr Ina Schaefer
Prof. Dr Wolgang J. Paul

Date: 22/03/2011

Signature(s):

1. OVERALL ASSESSMENT

a. Executive summary

[Assessment of the project highlighting the scientific/technical achievements of the project]

☐ Excellent progress (the project has fully achieved its objectives and technical goals for the period and has even exceeded expectations).

X Good progress (the project has achieved most of its objectives and technical goals for the period with relatively minor deviations).

☐ Unsatisfactory progress (the project has failed to achieve critical objectives and/or is not at all on schedule).

b. Main Recommendations

[Recommendations, e.g. on overall modifications, corrective actions at WP level, re-focusing of resources ...).

We make three main recommendations:

1) *We acknowledge that the choice of the MSC-51 processor allows obtaining exact complexity results. However, in order not to limit the results that can be obtained in the project to processors with very old architecture layouts from the WCET field, the project partners should explicitly consider how the assumptions on the hardware architecture influence the results obtained and in how far these assumptions could be a threat to the validity and generality of the results.*
   In order to allow the application of the proposed labelling approach for more modern processor architectures, we recommend to adapt the labelling approach such that basic complexity annotations can be obtained for program pieces of larger granularity than it is done now. This will allow transferring the results also to processors with pipelines and caches and enable the use of WCET tools to infer time bounds for the complexity annotations.

2) *The authors should quickly outline paper-and-pencil correctness proofs for each of the 7 stages of the CerCo compiler in order to allow for an estimation of the complexity and time required to complete the formalization of the proofs. We conjectured that the proof for the preservation of complexity bounds will not become more complex than in the toy compiler, because in no stage optimization across basic blocks are performed.*

3) *Based on the outline of the paper-and-pencil proofs, the authors should estimate the effort for two possible scenarios: a) proceed as planned with the MATITA system (this involves porting all proofs of the CompCert project to MATITA) or b) modify the existing proofs in Coq (this involves only to port results from D3.1 and D4.1 to*

Coq). Based on these estimates they should proceed in order to complete the project objective in duration of the project. If the project partners stick to the usage of MATITA, they should argue convincingly why this is a suitable choice and in how far MATITA is superior to Coq which is a widely used and well-accepted proof assistant.

2 OBJECTIVES and WORKPLAN

Please elaborate on the following issues:

a. Have the objectives for the period been achieved? In particular, has the project as a whole been making satisfactory progress in relation to the Description of Work (Annex I to the grant agreement)?

The objectives of the project as stated in the DoW out have been achieved in the first year of the project. However, we expect that the objectives for the next two years will be significantly harder to achieve. There are two main reasons for this:

i) Much of this project is based on work done in the CompCert project by Leroy et al. where the functional correctness of an optimizing compiler has already been formally shown. During the project, these results must be transferred from the Coq proof assistant to the MATITA system. It is clear that this will cost very much effort. It is much less clear how the goals of the project will profit by doing this. On the pure functional correctness side in the coming two years of the project, 6 intermediate languages and their semantics have to be defined. Then, formal simulation proofs between adjacent pairs of layers have to be constructed. With 6 intermediate languages, a source language and a target language, this makes 7 substantial formal proofs. This is very much work all of which exist already in Coq. The project partners should make very clear in how far this effort pays off for the results of the project.

ii) The choice of the MSC-51 processor and ignorance of results from the Worst Case Execution Time (WCET) field will severely limit the industrial interest in the project, even if all project goals are achieved. In contrast to modern processors, the MSC-51 has neither a pipeline nor caches. However, different memory regions of the MSC-51 processor apparently can be used as a software-controlled cache. This should be exploited in the project and made explicit as an achievement if the software-controlled cache can be handled by the results of the project.

Precise back annotations for such a simple processor can indeed be obtained on the basic block level as is done in the CerCo project. But with caches, one would always have to assume cache misses; in case cache hits and a two level cache hierarchy, one would overestimate memory access times by a factor of 100. This, of course, would make the estimates obtained by the approach of the CerCo project of very limited use.

In order to apply the labelling approach with state-of-the-art WCET analysis techniques for modern processor architectures, the granularity of program pieces for the basic annotations has to be much coarser. In a nutshell, the program pieces

must have such a large runtime that the initial cache and pipeline state has only a very small relative effect on their runtime behavior.

Because this larger program pieces 'only' affect the base case of the recursive analysis, the projects compiler can easily be changed to start with large program chunks; obviously, this comes at the expense that the estimates for the basic chunks can presently not be formally verified since they must be derived from results of WCET analysis tools.

b. Has each work package (WP) been making satisfactory progress in relation to the Description of Work (Annex I of the grant agreement)?

Each work package has made satisfactory progress with respect to the goals to be mentioned in the DoW. However, also for the results of the work packages the above mentioned problems hold. The choice of the MSC-51 processor as a target platform may severely limit industrial interest in the project's results. The usage of the MATITA system will require much additional effort for translating the existing results from the Coq system without obvious benefit for the CerCo project.

c Are the objectives for the coming period(s) i) still relevant and ii) still achievable within the time and resources available to the project? Is the approach and methodology still appropriate?

The objectives for the coming period are still relevant. However, we strongly recommend to consider seriously in how far the results obtained for the MSC-51 can be transferred to a modern processor architecture, along the lines describes above. We recommend that the results are generalized to other, more modern architectures in order to increase the impact of the project, in particular with respect to industrial applicability. Further, the usage of the MATITA system should be evaluated whether the effort spent to transfer an existing Coq formalization to MATITA can be justified with the benefits obtained by the usage of the MATITA system.

d. Are the planned deliverables for the reporting period approved? (please fill in the list of deliverables due in this reporting period)

DELIVERABLES LIST STATUS			
No.	**Title**	**Status (Approved/Rejected)**	**Remarks**
D1.1	Periodic Activity Report and Financial Statements	approved	
D2.1	Compiler Design and intermediate languages	approved with addendum	
D2.2	Untrusted Cost-annotating OCaml Compiler	approved	
D3.1	Executable Formal Semantics of C	approved	
D4.1	Executable Formal Semantics of Machine Code	approved	
D6.1	Project Web Site and Software Repository	approved	
D6.2	Plan for the Use and dissemination of foreground	approved with addendum	

Detailed Assessment of Deliverables:

- D1.1: approved

- D2.1: The deliverable describes the intermediate layers of an optimizing compiler as used by Leroy et al in a very generic way. The difference between results that are used form the CompCert project and the novel contributions made by the CerCo project are not apparent.
  Only very few lines of the deliverable are devoted to the optimizations performed by the implemented compiler, although optimizations are explicitly mentioned in the DoW. This description of the optimizations needs to be improved.
  References to related work concerning general approaches for invariant generation for termination analysis, general approaches for resource analysis and for the usage of the labelling approach for analysing intentional program properties are missing entirely. In particular, the authors need to consider the following reference:

  - C. Ferdinand, R. Heckmann, T. Le Sergent, D. Lopes, B. Martin, X. Fornari, F. Martin: Combining a High-Level Design Tool for Safety-Critical Systems with a Tool for WCET Analysis on Executables, available from http://www.esterel-technologies.com/technology/WhitePapers/
  - Elvira Albert, Puri Arenas, Samir Genaim, Miguel Gómez-Zamalloa, German Puebla, Diana V. Ramírez-Deantes, G. Román, Damiano Zanardini: Termination and Cost Analysis with COSTA and its User Interfaces. Electr. Notes Theor. Comput. Sci. 258(1): 109-121 (2009)

  Also, annotations for space consumption have been done at least twice in the formal verification community:

  - Gilles Barthe, Mariela Pavlova, and Gerardo Schneider. 2005. Precise Analysis of Memory Consumption using Program Logics. In *Proceedings of the Third IEEE International Conference on Software Engineering and Formal Methods* (SEFM '05). IEEE Computer Society, Washington, DC, USA, 86-95.
  - Alkassar et al: Balancing the load: Leveraging semantic stack for system verification. In Klein et al (eds.). Journal of Automated Reasoning. Special Issue on Operating System Verification. Pages 389-454, 2009

  The deliverable is accepted under the condition that an addendum providing a survey of related approach is provided by end of April 2011.

  The main properties of the compiler are detailed in the deliverable. They should be useful for the certified compiler that will be developped in MATITA. The deliverable should explain 1) why these properties have been proved in Coq and not in MATITA, and 2) why two of these proofs are only paper and pencil proofs.

- D2.2: This deliverable is also a fairly generic report on compiler construction with too few detail on cost estimate generation. One would like to learn more about the times required to access the different memory regions of the MSC-51. As we learned in the presentation at the review meeting, cost estimation for arithmetic operations heavily depends on operands, because the MSC-51 has only an 8-bit ALU. Dealing

with this restriction requires much effort and solves a problem not existing in modern machines which have wider ALUs. It seems advisable to replace these estimates by worst case estimates. With this, one will obviously sacrifice the ability for precise cost computations. Some features of the compiler are not yet implemented due to the switch from the toy compiler to MIPS described in D2.1 to the compiler for the MSC-51 platform in D2.2.
The C#minor language of the CompCert compiler is not an intermediate language of the CerCo prototype compiler. If this means that it will not be an intermediate language of the CerCo certified compiler, then the proof of correctness of the translation from Clight to Cminor will be more tedious.
Function pointers are not supported but the deliverable states that they could be supported with a minimum effort. This is confusing.

- D3.1: This deliverable is a generic report on C semantics with little detail about the consequences of the specific choice of processor. One would have liked to see details how the 'usual' semantics are affected by the fact, that variable declarations contain compiler directives indicating in which of the 3 memory regions of the MSC-51 the variable is to be allocated. This is substantial work which would have been much easier if the existing definitions in Coq would have been used and modified to deal with the three memory regions.
  The example of a simple Clight program written as a MATITA term is difficult to read, mainly because of all the details given in the abstract syntax tree. The benefit of generating such MATITA terms should be explained in the deliverable.

- D4.1: This deliverable essentially describes an MSC-51 emulator written in the functional language OCaml. Obviously, this not completely easy for CISC processors.
  An emulator written in MATITA has been derived from the OCaml emulator. It relies on the development of a small library defining basic data structures and also on dependent types. It would be interesting to know if this library already exists in Coq and to explain why the use of dependent types in this development would differ from a Coq version of this development.

- D6.1: approved

- D6.2: In this deliverable, the dissemination activities in academia and in industry in the previous project period are reported. A detailed plan for future dissemination activities in the following two project years is missing. The deliverable is approved subject to an addendum describing future plans to be delivered by end of April 2011.
  With respect to dissemination in academia, at least the following three relevant conferences are missing: VSTTE, SSV and FMCAD.
  For dissemination in academia, we propose to organize a satellite workshop at one of the major conferences.
  For dissemination in industry, we propose to propose a tutorial on the CerCo compiler at an academic conference with a large number of participants from industry. Alternatively, the project can participate in a research project symposium, such as the one to be held at ECOOP 2011.

Minor comment:

The authors of the different deliverables should decide on a uniform spelling of frequently used terms, such as OCaml, Clight, etc.

3. RESOURCES

To the best of your estimate, have resources used, i.e. personnel resources and other major cost items, been (i) utilised for achieving the progress, (ii) in a manner consistent with the principle of economy, efficiency and effectiveness? If applicable, please comment on large deviations with respect to the planned resources.

The resources have been used appropriately. However, we have heard that the local administration in Bologna has made it impossible to fill positions financed by EU money. If true, this seems to be an incredibly unproductive thing to do.

4. IMPLEMENTATION OF THE PROJECT

a. Has the project management - with respect to technical, administrative and financial management - been performed as required?

So far yes. But formal proofs obtained by different teams have not been integrated yet. A software repository in SVN has been established and includes – as we have learned in the meeting – also the formal definitions and theories. The project management should be aware of the fact that SVN permits to

- *make temporary modifications of definitions of theories*
- *work on these for a while and then*
- *forget to synchronise the definitions theories with the originals used elsewhere in the field*

Therefore, we advise the management to periodically check in the future that identical definitions are use everywhere in the projects.

b. Has the collaboration between the beneficiaries been effective?

So far yes.

c. Do you identify evidence of underperforming beneficiaries, lack of commitment or change of interest of any beneficiaries?

No.

Comments:

5. USE AND DISSEMINATION

a. Is the project likely to have scientific, technical, commercial, social, or environmental impact. Is the plan for the use of foreground?

If caches and pipelines of modern processor architectures are continued to be ignored, the impact of the project might very well turn out to be disappointing, since the results will only be applicable to a very limited number of ancient processors.

b. Are there any results or success stories that merit publicity?

So far no.

c. Have the beneficiaries disseminated project results and information adequately (publications, conferences, website, inclusion of stakeholders…)?

So far no, but this is only the first year of the project. Several papers have been submitted and are currently under review.

6. OTHER ISSUES

a. Have policy-related and/or regulatory issues been properly handled (if applicable)?
b. Have ethical issues been appropriately handled (if applicable)?
c. Have safety issues been properly handled (if applicable)?
d. Has progress on Gender Equality Actions been satisfactory (if applicable for this reporting period)?

Comments: no irregularities observed