Thesis Advisor Position

Thesis: Garbage Collection in Software Performance Engineering
Author: Peter Libič
Advisor: Petr Tůma
Submitted: August 2015

In the 2015 ranking of the most popular programming languages by IEEE, only two languages out of the top ten use explicit memory management. It is safe to say that some form of garbage collection has become a common feature of most runtime environments – in fact, so common that application developers can often work with almost no knowledge of the garbage collection mechanism their environment uses. There is a downside to this – garbage collection can introduce significant performance overhead that is not transparent from the developer perspective, and implementing applications with reasonable performance behavior becomes a matter of trial and error.

Obviously, the question of garbage collector performance is studied intensively by the garbage collection research community, however, their insights tend to concern technical intricacies of the garbage collection implementation, which may simply be too distant from the common developer concerns to be useful for software performance engineering. The work of Peter Libič addresses the problem by looking at garbage collection from the performance engineering perspective – it shows how the garbage collection overhead relates to certain basic workload patterns, what information is relevant to estimating garbage collection frequency and duration, and what changes in the overhead can be expected with certain changes to application allocation behavior.

Department of Distributed and Dependable Systems
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
phone +420-221914245, fax +420-221914323
http://d3s.mff.cuni.cz

The thesis does not provide a simple solution that a software performance engineer can apply to manage the garbage collection overhead. Instead, it examines the relevant concerns and their complex interactions, demonstrating that although we can sometimes predict the overhead, we are still far away from integrating garbage collection behavior into regular performance engineering processes. In light of the programming language statistics above, this is indeed worrying.

There are two points of the thesis that I would like to emphasize. One is the size of the experiments, which is way above the typical measurements done in software performance engineering. Peter had to work hard to overcome the related issues – long experiment runtimes, reduced space for interactively correcting mistakes, large data processing times, and so on. The other point is the inherent difficulty of the task – mainstream garbage collectors were perhaps designed to be efficient, but certainly not to be predictable, and definitely not to be simple. To pursue a thesis topic that hinges on finding regularities in the complex and fragile garbage collector behavior requires dedication that Peter demonstrated in the extreme.

To conclude, I believe the thesis represents a robust research work that advances a highly difficult but practically relevant topic. I am glad to recommend that Peter Libič be awarded the doctoral degree.

Petr Tůma

Department of Distributed and Dependable Systems
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
phone +420-221914245, fax +420-221914323
http://d3s.mff.cuni.cz