Abstract

Dependability benchmarks are designed to assess, by quantifying through quantitative performance and dependability attributes, the behavior of systems in presence of faults. In this type of benchmarks, where systems are assessed in presence of perturbations, not being able to select the most suitable system may have serious implications (economical, reputation or even lost of lives). For that reason, dependability benchmarks are expected to meet certain properties, such as *non-intrusiveness*, *representativeness*, *repeatability* or *reproducibility*, that guarantee the robustness and accuracy of their process. However, despite the importance that comparing systems or components has, there is a problem present in the field of dependability benchmarking regarding the analysis and comparison of results.

While the main focus in this field of research has been on developing and improving experimental procedures to obtain the required measures in presence of faults, the processes involving the analysis and comparison of results were mostly unattended. This has caused many works in this field to analyze and compare results of different systems in an ambiguous way, as the process followed in the analysis is based on argumentation, or not even present. Hence, under these circumstances, benchmark users will have it difficult to use these benchmarks and compare their results with those from others. Therefore extending the application of these dependability benchmarks and perform cross-exploitation of results among works is not likely to happen.

This thesis has focused on developing a methodology to assist dependability benchmark performers to tackle the problems present in the analysis and comparison of results of dependability benchmarks. Designed to guarantee the fulfillment of dependability benchmark's properties, this methodology seamlessly integrates the process of analysis of re-

sults within the procedural flow of a dependability benchmark. Inspired on procedures taken from the field of operational research, this methodology provides evaluators with the means not only to make their process of analysis explicit to anyone, but also more representative for the context being.

The results obtained from the application of this methodology to several case studies in different domains, will show the actual contributions of this work to **improving the process of analysis and comparison of results in dependability benchmarking for computer systems**.