A Preliminary Framework for Dependability Benchmarking

Karama Kanoun, Henrique Madeira, Jean Arlat

A Preliminary Framework for Dependability Benchmarking

Karama Kanoun[*], Henrique Madeira[**] and Jean Arlat[*]

[*] LAAS-CNRS, 7, Avenue du Colonel Roche, 31077 Toulouse Cedex 4 — France

[**] DEI-FCTUC, University of Coimbra, 3030 Coimbra — Portugal

Contact author
Karama Kanoun
LAAS-CNRS, 7, Avenue du Colonel Roche
31077 Toulouse Cedex-4, France
kanoun@laas.fr

Word count: 1255

Abstract

This paper outlines a preliminary framework for defining dependability benchmarks of computer systems that is being investigated by the European project DBench. The multiple dimensions of the problem are classified, then examples of benchmarking scenarios are presented. Finally, some research issues are discussed.

Keywords

Dependability assessment, dependability benchmark

A Preliminary Framework for Dependability Benchmarking

Karama Kanoun[*], Henrique Madeira[**] and Jean Arlat[*]
[*] LAAS-CNRS, 7, Avenue du Colonel Roche, 31077 Toulouse Cedex 4 — France
[**] DEI-FCTUC, University of Coimbra, 3030 Coimbra — Portugal
(kanoun@laas.fr, henrique@dei.uc.pt, arlat@laas.fr)

Abstract

This paper outlines a preliminary framework for defining dependability benchmarks of computer systems that is being investigated by the European project DBench[*]. The multiple dimensions of the problem are classified, then examples of benchmarking scenarios are presented. Finally, some research issues are discussed.

1. Introduction

The goal of benchmarking the dependability of computer systems is to provide generic ways for characterizing their behavior in the presence of faults. Benchmarking must provide a uniform, and repeatable way for dependability characterization. The key aspect that distinguishes benchmarking from existing evaluation and validation techniques is that a benchmark fundamentally should represent an agreement that is accepted by the computer industry and by the user community. This technical agreement should state the measures, the way these measures are obtained, and the domain in which these measures are valid and meaningful. A benchmark must be as representative as possible of a domain. The objective is to find a representation that captures the essential elements of the domain and provides practical ways to characterize the computer dependability to help system manufacturers and integrators improving their products and end-users in their purchase decisions.

The DBench project aims at defining a conceptual framework and an experimental environment for dependability benchmarking. This paper summarizes our first thoughts on the conceptual framework, as investigated in [1]. Section 2 identifies the various dimensions of dependability benchmarks. Section 3 presents some benchmarking scenarios. Section 4 introduces some research issues.

2. Benchmark Dimensions

The definition of a framework for dependability benchmarking requires first of all the identification and the clear understanding of all impacting dimensions. The latter have been grouped into three classes as shown in Figure 1.

- *Categorization dimensions* allow organization of the dependability benchmark space into different categories.
- *Measure dimensions* identify dependability benchmarking measure(s) to be assessed depending on the choices made for the categorization dimensions.
- *Experimental dimensions* include all aspects related to experimentation on the target system to get the base data needed to obtain the selected measures.

Categorization dimensions

The *target system nature and application area* impact at the same time measures to be evaluated and measurements to be performed on the target system to obtain them.

Benchmark usage includes:

- *Benchmark performer*: Person or entity who performs the benchmark (e.g., manufacturer, integrator, third party, end-user). These entities have i) different visions of the target system, ii) distinct accessibility as well as observability levels for experimentation, and iii) different expectations from the measures.
- *Benchmark user:* person or entity who actually uses the benchmark results, who could be different from the benchmark performer.
- *Benchmark purpose*: results can be used either to characterize system dependability capabilities in a qualitative manner, to assess quantitatively these capabilities, to identify weak points or to compare alternative systems.
- *Results scope*: Benchmark results can be used internally or externally. External use involves standard results that fully comply with the benchmark specification. for public distribution, while internal use means that it is used for system validation and tuning.

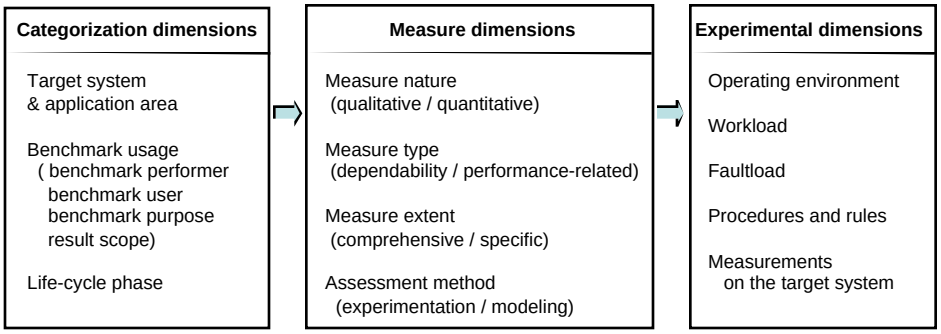

Figure 1 - Dependability benchmarking dimensions

[*] IST 2000-25425 http://www.laas.fr/DBench

Measure dimensions

The various usage perspectives impact the type (and the detail) of benchmark measures. Typically, end-users are interested in dependability measures defined with respect to the expected services (i.e., comprehensive measures, such as availability), while manufacturers and integrators could be more interested in specific measures related to particular features of the target system (e.g., error detection coverage). Comprehensive measure may be evaluated based on modeling.

Experimentation dimensions

The operating environment traditionally affects very much system dependability. The workload should represent a typical operational profile for the considered application area. The faultload consists of a set of faults and exceptional conditions that are intended to emulate the real exceptional situations the system would experience. This is clearly dependent on the operating environment for the external faults, which in turn depends on the application area. Internal faults (e.g., software and some hardware faults) are mainly determined by the actual target system implementation. A dependability benchmark must include standards for conducting experiments and to ensure uniform conditions for measurement. These standards and rules must guide all the processes of producing dependability measures using a dependability benchmark.

3. Benchmark Scenarios

The set of successive steps for benchmarking dependability together with their interactions form a benchmarking scenario. Benchmarking starts by an analysis step for allocation of specific choices to all categorization and measure dimensions. The selection of the experimental dimensions is then achieved based according to theses choices.

Figure 2 gives a high level overview of the activities and their interrelations (represented by arrows A to E) for system dependability benchmarking. To illustrate how this general framework can be used in real situations, we have selected four examples of benchmark scenarios (S1 to S4).

S1: Benchmark based on experimentation only

S1 includes analysis and experimental steps, and link A. It is actually an extension of the well-established performance benchmark setting.

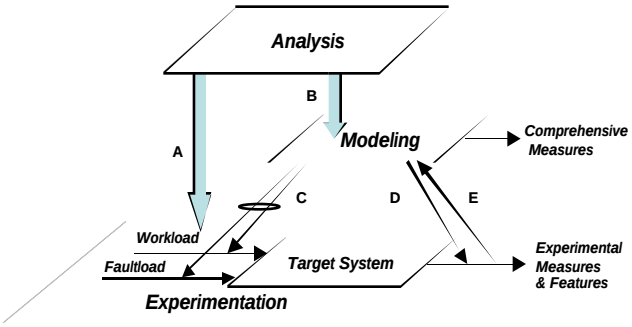

Figure 2 - Dependability benchmarking scenarios

S2: Experimentation supported by modeling

S2 includes the three steps and links A, B, C and D. The experimentation is guided, at least partially, by modeling.

S3: Modeling supported by experimentation

S3 includes the three steps and links A, B and E. Experimentation supports model validation and refinement. The expected outputs are comprehensive measures obtained from processing the model(s). However, the experimental measures and features, assessed for supporting modeling, may be made available from the experiments.

S4: Modeling and experimentation

S4 is a combination of S2 and S3 and includes all steps and all links. Its outputs are experimental measures and features, and comprehensive measures based on modeling.

4. Some Research Issues

Representativeness is a crucial concern, as benchmark results must characterize the addressed aspects of the target system in a realistic way. Regarding established performance benchmarks, the problem is reduced to the representativeness of performance measures and of the workload. For dependability, it is also necessary to define representative dependability measures and representative faultloads. Although the problem seems clearly more complex than for performance benchmarks, the pragmatic approach used in the established performance benchmarks offers a basis for identifying adequate solutions for dependability benchmarking representativeness.

It is worth mentioning that many technical problems still need to be resolved. The subsequent points summarize crucial research issues.

- The adoption of the workloads of established performance benchmarks is the starting point for the definition of workloads. However, some work still has to be done. For example one has to check whether the way the application spectrum as partitioned by the performance benchmarks is adequate for this new class of dependability benchmarks.
- The definition of representative faultloads encompasses specific problems that are currently being studied
- Definition of meaningful measures. In particular, special attention should be paid to confidence and accuracy of measurements and the possible impact of the measurements on the target system behavior (intrusiveness).

Finally, dependability benchmarks must meet certain properties to be valid and useful. One key property is reproducibility. In fact, benchmarks can be accepted only if results can be repeated and reproduced by another party.

Acknowledgement

All partners of the DBench project contributed to the work presented in this paper. The list of names is too long to be included in a two-page position paper.

Reference

[1] H. Madeira, K. Kanoun, J.Arlat, Y. Crouzet, A. Johanson, R Lindström, "Preliminary Dependability Benchmark Framework", DBench deliverable, September 2001. Available at http://www.laas.fr/dbench/delivrables.html