From Performance to Dependability Benchmarking:

A Mandatory Path

TPC TC 2009

Marco Vieira, Henrique Madeira

University of Coimbra – Portugal

mvieira@dei.uc.pt, henrique@dei.uc.pt

I'm here to challenge you!!!

- **Is computer benchmarking only about performance?**

  - **NO!!!**

- **E.g. nowadays most systems need to guarantee high availability and reliability**

  - It is mandatory to shift the focus from measuring performance to the measurement of both performance and dependability

- **Don't computers fail?**

  - What is the impact of failures into the system?

Why do computers fail?

Too many reasons…

- Hardware problems

Too many reasons…

- Environment problems

Too many reasons…

- Bad configuration

Too many reasons…

■ Misuse

Too many reasons…

- ■ Not proven design

But most of the times it is due to…

- Software problems

What can we do?

- Don't use computers ☺
  - Sooner or later they will fail!!!

- Built better systems
  - Many have tried... most have failed ☺

- But... how do we know if a system is more dependable than another?

Guess!

Not a good answer…

- Prophets do not exist!

- How to improve this?

  - That is my challenge for you:

    Dependability Benchmarking

Outline

- **Performance benchmarking**
  - A Well Established Field

- **Dependability benchmarking**
  - A Recent Research Field

- **Dependability benchmark example**
  - DBench-OLTP

- **Research needs and challenges**

Performance benchmarking

Standard procedures and tools to evaluate and compare systems or components according to specific performance measures

- Started four decades ago
  - Whetstone programs (1972)
    - Instructions executed per second

- Two key organizations working on the performance benchmarking business
  - TPC (Transaction Processing Performance Council)
  - SPEC (Standard Performance Evaluation Corporation)

Dependability benchmarking

Procedures to measure both the dependability and performance of systems or components

- Compare systems or components from a dependability point-of-view:
  - Availability
  - Reliability
  - Safety
  - Confidentiality
  - Integrity
  - Maintainability
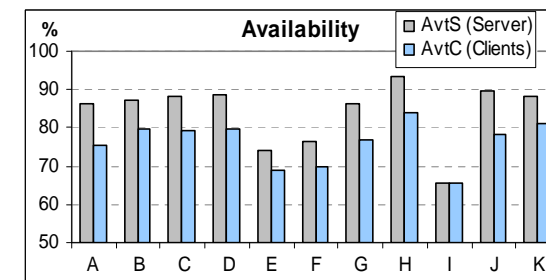
Components of a dependability benchmark

Performance Benchmark

Workload
Performance Measures

+

Faultload
Dependability Measures

=

Dependability Benchmark

Workload
Faultload
Performance & Dependability Measures
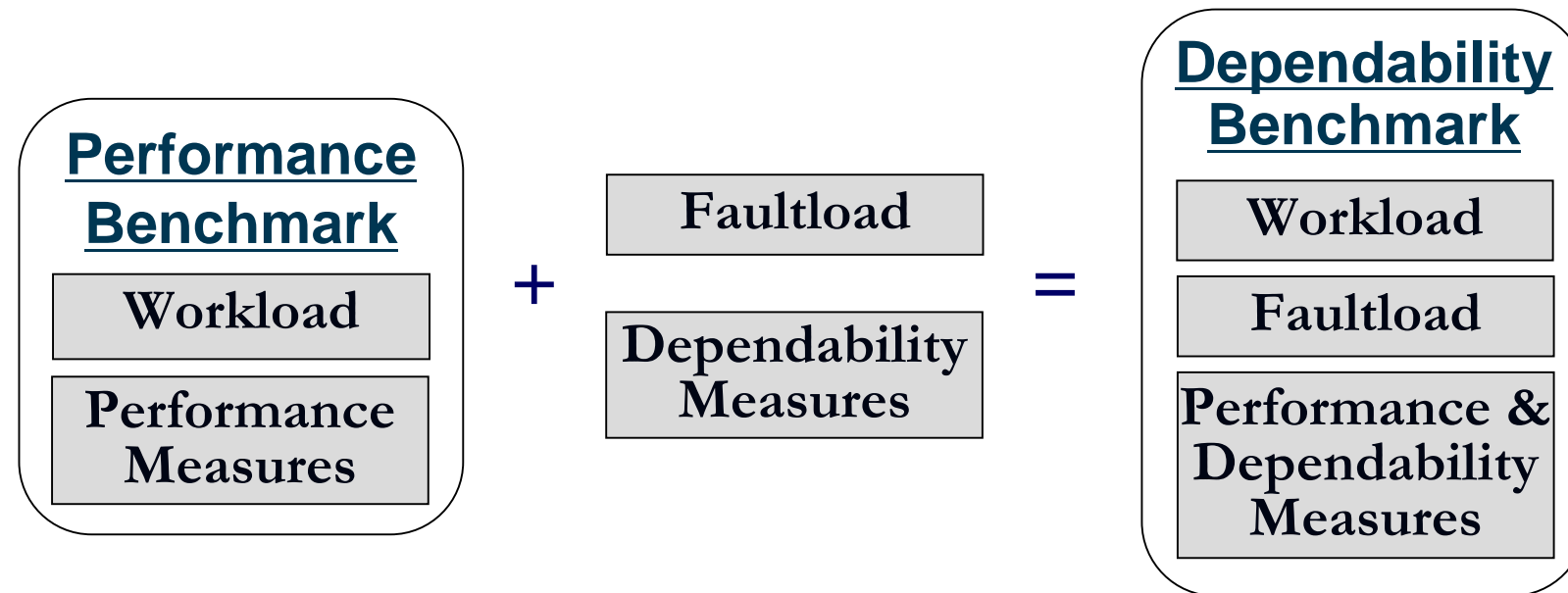
- Procedure and rules

- Experimental setup

Work on dependability benchmarking… (1)

- **Special Interest Group on Dependability Benchmarking (SIGDeB)**
  - Created by the International Federation for Information Processing (IFIP) Working Group 10.4 in 1999
  - Promote the research, practice, and adoption of benchmarks for computer systems dependability

- **Assessing, Measuring and Benchmarking Resilience (AMBER)**
  - Coordination Action supported by the European Commission in the 7th Framework Programme
  - Coordinating and advancing research in resilience measurement and benchmarking

Work on dependability benchmarking... (2)

- ## The DBench Project

  - Funded by the European Commission, under the Information Society Technologies Programme (IST), FP5

  - Devise benchmarks to evaluate and compare the dependability of COTS and COTS-based systems
    - Embedded, real time, and transactional systems

  - Several proposals:
    - General purpose operating systems
    - Real time kernels for onboard space systems
    - Engine control applications for automotive systems
    - On-line transaction processing systems
    - Web-servers

Work on dependability benchmarking… (3)

- ## Berkeley University

  - Benchmarking the dependability of human-assisted recovery processes

- ## Sun Microsystems

  - Extremely active on dependability benchmarking!!!

  - High-level framework specifically dedicated to availability benchmarking of computer systems

  - Within this framework, two specific benchmarks have already been developed
    - System's robustness (degree of protection against outage events) in handling maintenance events
    - System recovery in a nonclustered standalone system

Work on dependability benchmarking... (4)

- ## Intel Corporation
  - Focused on benchmarking semiconductor technology
  - E.g. impact of semiconductor technology scaling on neutron induced SER (soft error rate)

- ## IBM Autonomic Computing Initiative
  - Quantify a system's level of autonomic capability
    - Defined as the capacity of the system to react autonomously to problems and changes in the environment
  - The goal is to produce a suite of benchmarks covering the four autonomic capabilities
    - Self-configuration, self-healing, self-optimization, and self-protection

Benchmark example: Dbench-OLTP

- Compare db-centric transactional systems

- Follows the style of the TPC benchmarks
  - (Customized) TPC-C workload

- Structured in Clauses:
  - Clause 1 – Preamble
  - Clause 2 – Benchmark Setup
  - Clause 3 – Benchmarking Procedure
  - Clause 4 – Measures
  - Clause 5 – Faultload
  - Clause 6 – Full Disclosure Report

Clause 4 – Measures

- **Baseline performance measures:**
  - tpmC: Number of transactions executed per minute
  - $/tpmC: Price per transaction

- **Performance measures in the presence of faults:**
  - Tf: Transactions executed per minute with faults
  - $/Tf: Price-per-transaction in the presence of faults

- **Dependability measures:**
  - Ne: Number of data errors detected
  - AvtS: Availability from the server point-of-view
 - AvtC: Availability from the end-users point-of-view

Clause 5 – Faultload

- **Three different faultloads can be considered**

- **Operator faults**
  - Emulate database administrator mistakes

- **Software faults**
  - Emulate software bugs in the operating system

- **High-level Hardware failures**
  - Emulates hardware component failures, such as:
    - Power failures
    - Disk failures
    - Corruption of storage media

Benchmarking example

System	Hardware	Operating System	DBMS	DBMS Config.
A	• *Processor*: Intel Pentium III 800 MHz • *Memory:* 256MB • *Hard Disks:* Four 20GB/7200 rpm • *Network:* Fast Ethernet	Win2k Prof . SP 3	DB-1	Conf. A
B		Win2k Prof . SP 3	DB-2	Conf. A
C		WinXp Prof. SP 1	DB-1	Conf. A
D		WinXp Prof. SP 1	DB-2	Conf. A
E		Win2k Prof . SP 3	DB-1	Conf. B
F		Win2k Prof . SP 3	DB-2	Conf. B
G		SuSE Linux 7.3	DB-1	Conf. A
H		SuSE Linux 7.3	DB-2	Conf. A
I	• *Processor:* Intel Pentium IV 2 GHz • *Memory:* 512MB • *Hard Disks*: Four 20GB / 7200 rpm • *Network:* Fast Ethernet	Win2k Prof . SP 3	DB-1	Conf. A
J		Win2k Prof . SP 3	DB-2	Conf. A

DBench-OLTP Results summary
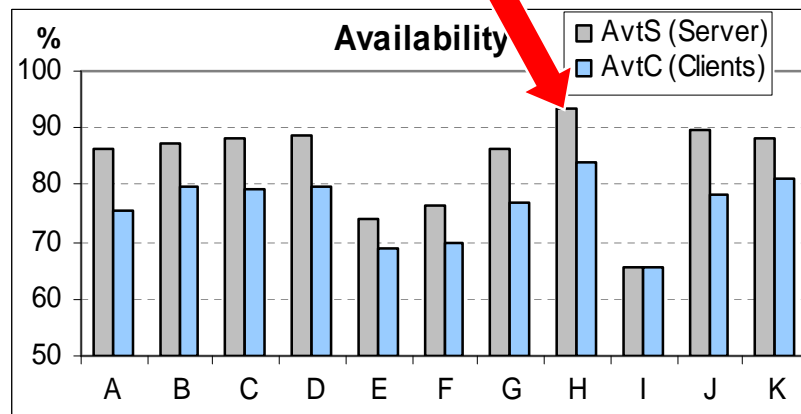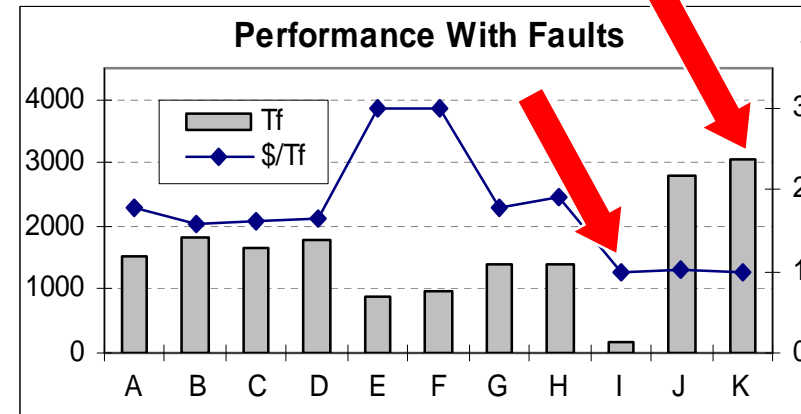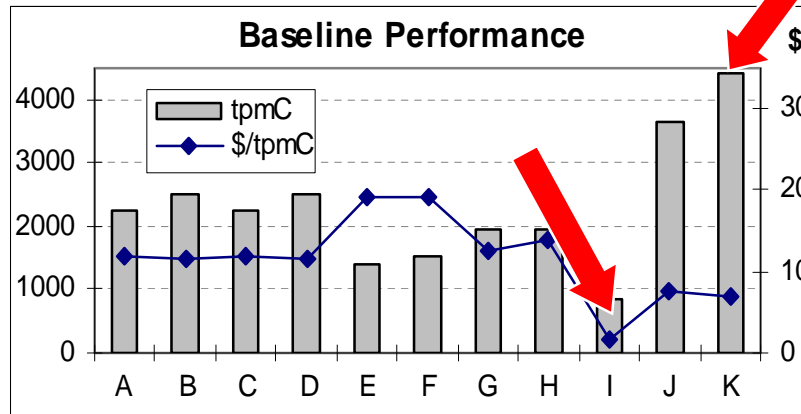
- Performance
- Availability
- Price

Vieira, M., Madeira, H.: A Dependability Benchmark for OLTP Application Environments, 29th Intl Conference on Very Large Data Bases, VLDB 2003, 2003

Research needs and challenges

- Many works have been conducted in the area of dependability benchmarking

- Some key issues remain open:
  - Availability of agreed dependability benchmarking measures
  - Elaboration of adaptable benchmarking processes
  - Development of benchmarking frameworks
  - Integration of benchmarking with the design methodologies
  - Maintenance to avoid negative effects (e.g., gamming)
 - **Support from a standardization body or industry**

Conclusions

- **Several works on dependability benchmarking have been conducted in the past**
    - At both academia and industry

- **No work has yet achieved the status of a real benchmark**
    - Endorsed by a standardization body or corporation

Dependability benchmarking is a mandatory path!!!

- **Isn't it time for TPC to start paying attention to it?** ☺

Bibliography
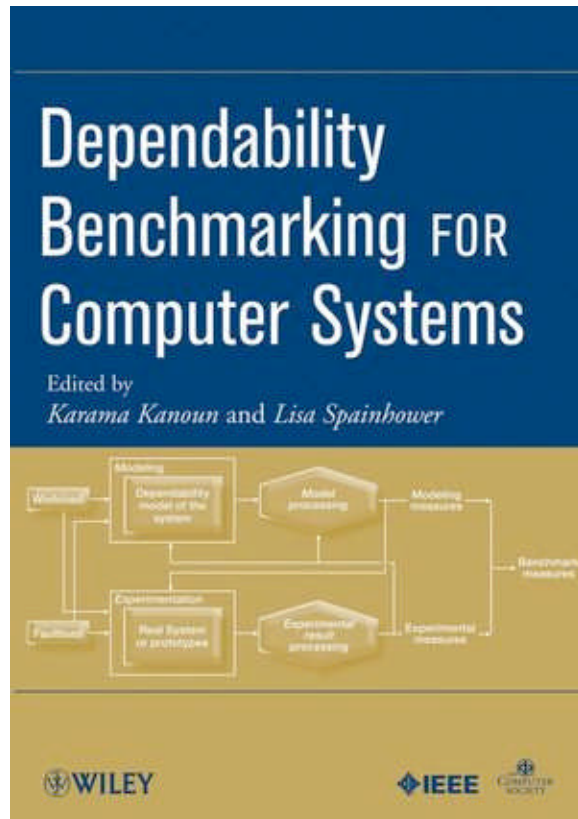
Dependability Benchmarking for Computer Systems
Karama Kanoun, Lisa Spainhower (Eds)
ISBN: 978-0-470-23055-8
Wiley-IEEE Computer Society Press
August 2008

Questions & Comments

Marco Vieira, Henrique Madeira

University of Coimbra – Portugal

mvieira@dei.uc.pt, *henrique@dei.uc.pt*