Middleware Benchmarking Project
Regression Benchmarking

Tomáš Kalibera
DISTRIBUTED SYSTEMS RESEARCH GROUP
http://nenya.ms.mff.cuni.cz

CHARLES UNIVERSITY PRAGUE
Faculty of Mathematics and Physics
Outline

• DSRG benchmarking history
• Motivation for regression benchmarking
• Problems for research
• Solutions
• Work in progress
• Desired regression benchmarking suite
• Future prospects
• Publications
History

1998-1999 Comparison of CORBA middleware for MLC Systeme (now Deutsche Post Com)

1999-2000 Comparison of EJB middleware for MLC Systeme (now Deutsche Post Com)
1999-2000 EJB Comparison Project (OSMOSE)

2000-2003 Testing of CORBA middleware for IONA Technologies and Borland International
2000-2004 Open CORBA Benchmarking Project
People

staff

Petr Tůma

phd students

Lubomír Bulej
Tomáš Kalibera

former members of the team

Adam Buble
Radek Pospíšil
Marek Procházka
Motivation: Quality Assurance of Performance

• Comparison benchmarks
  ▪ find out that a product has become slower over last few months of development
  ▪ forget it or start again

• Regression benchmarks
  ▪ find out that a product has become slower after incorporating recent changes
  ▪ fix it and go on
Problems for Research

• Benchmarking in general
  ▪ benchmark granularity (model application vs. feature specific benchmarks)
  ▪ do not change performance of the measured system by adding the measuring code

• Automated benchmarking
  ▪ (in)sufficient warm up
  ▪ when to stop measuring (how much data we need to get stable results)
Problems: High Variation in Measurements

- Different results even for repeated runs of a benchmark on the same system
  - interferences from the operating system and hardware
  - initial conditions of the process that are not changed during benchmark run
Solutions: Simple Benchmarks

• Handling of interferences
  ▪ using robust statistics (median)
  ▪ keeping measurement duration below the periodicity of operating system events (context switches)

• Handling of initial conditions
  ▪ running benchmarks multiple times
  ▪ comparing probability distributions of results of individual runs
Results: Regressions in Marshalling (TAO)
• Detect rare regressions in complex benchmarks
  ▪ complex benchmarks can reveal regressions in interplay of application parts
  ▪ results are hard to interpret
Work in Progress: Complex Benchmarks

- Apply cluster analysis on complex results
- Match clusters of measurements of the benchmarks
- Compare clusters of measurements by methods used for simple benchmarks
RUBiS: Effect of Uninitialized Random Seed

Original(R) and Corrected Random Seed
Client on Linux 2.4

Original(R) and Corrected Random seed
Client on SMP, Linux 2.6
Our Vision: Regression Benchmarking Suite

- Control Module
- Analysis Module
- Benchmark Framework
- Benchmark Modules
- System Under Test
- Notifications Repository
- Results Repository

Future Prospects

Tomáš Kalibera
ObjectWeb Benchmarking Workshop, Grenoble, 5 Apr 2004
Future Prospects

• Automatic analysis of results
  ▪ automatic detection of regressions
  ▪ support for both feature specific and application benchmarks

• Results repository
  ▪ efficient storage of huge data
  ▪ common interfaces
  ▪ support for statistical software

• Buble, A., Bulej, L., Tuma, P.: CORBA Benchmarking: A Course With Hidden Obstacles, in proceedings of the IPDPS Workshop on Performance Modeling, Evaluation and Optimization of Parallel and Distributed Systems (PMEOPDS 2003), Nice, France

• Tuma, P., Buble, A.: Overview of the CORBA Performance, in proceedings of the 2002 EurOpen.CZ Conference, Znojmo, Czech Republic


• Prochazka, M.: Advanced Transactions in Enterprise JavaBeans, Proceedings of EDO 2000, Davis (CA), USA, Springer Verlag, LNCS