Computer Science 703
Advance Computer Architecture
2010 Semester 1
Lecture Notes
27Apr10
Methods 2: Tools & Benchmarks

James Goodman

Department
of
Computer Science

Evaluating Simulation Results

- Results are counts of occurrences
  - How many clock cycles
  - How many arrivals
  - How many failures
  - How many ???
- Results may be repeatable
  - If inputs are repeatable
  - Have to introduce randomness!
  - Multiprocessors often introduce randomness
  - Small changes in input can produce large changes in results
- Multiple simulations may be required
  - Multiple parameters, multiple combinations
- Would like to simulate multiple cases simultaneously

Validation

- Can you believe the results?
  - Do you know how to interpret the results?
  - Do you understand what is being modelled?
  - Do you understand how it is being modelled?
- Results can mislead regarding accuracy
  - Exact, repeatable counts are precise
  - Accuracy depends on assumptions
- Studying results: sanity checks
  - Compare against real system (or similar real system)
  - Compare small variations in parameter(s)
  - Compare against other methods
    - Other simulators
    - Analytical models
  - Experience is important

Simulators are Everywhere

Many simulators are
- Hard to use
- Poorly documented
- Poorly specified
- Unsupported

Good simulators may be free (to academics) but offered "as-is"

All but simplest simulators require substantial experience/understanding to interpret results

How to Report Performance

- Clock rate?
- MIPS? MFLOPS?
- Peak performance?
- Time to execute a programme?
  - What programme?

Quotations from my Father

"Figures don't lie, but liars do figure!"

Programme Benchmarks

- What is an appropriate programme?
  - Choice of programme can dramatically affect results
    - Computation limited (integer or floating point?)
    - Memory limited
    - Control limited
  - Changes in some parameters can dramatically affect results

What is a "good" benchmark?

- What are we trying to measure
  - Computation limited (integer or floating point?)
  - Memory limited
  - Control limited
- What is an appropriate programme?
  - Choice of programme can dramatically affect results
  - Parameters matter
    - Singly or collectively parameters can dramatically affect results

Combining Multiple Benchmarks

- J.E. Smith, "Characterizing computer performance with a single number," *CACM*, v. 31, #10 (October 1988), pp. 1202-1206.
  - Arithmetic mean (of MFLOPS)
  - Geometric mean
  - Harmonic mean

"... the time required to perform a specified amount of computation is the ultimate measure of computer performance."

Property 1. A single-number performance measure for a set of benchmarks expressed in units of time should be directly proportional to the total (weighted) time consumed by the benchmarks.

Property 2. A single-number performance measure for benchmarks expressed as a rate should be inversely proportional to the total (weighted) time consumed by the benchmarks.