

Challenges In Capturing Real World Workloads In Benchmarks

Summary Thoughts/Opinions

October 25, 2004

7th Annual IEEE Workshop on Workload Characterization

Developed by

Jeff Reilly

Intel, Principal Engineer

SPEC, Chair of SPEC CPU Subcommittee/Member of Board of Directors

“What has the power to generate heated controversy, hard feelings, and bold accusations? Sex scandal? Litigation? Nope – try benchmarks.” –

<http://www.arstechnica.com>, J. Stokes

A GOOD Benchmark Is...

- Relevant
- Recognized
- Simple
- Portable
- Scalable



Source: Jim Gray: Stanford Mini-Course on Database Performance

Not all benchmarks (including some popular ones) have all of these characteristics! Question their usefulness...

Benchmark Challenges or “What Makes A Good Benchmark Go Bad?”

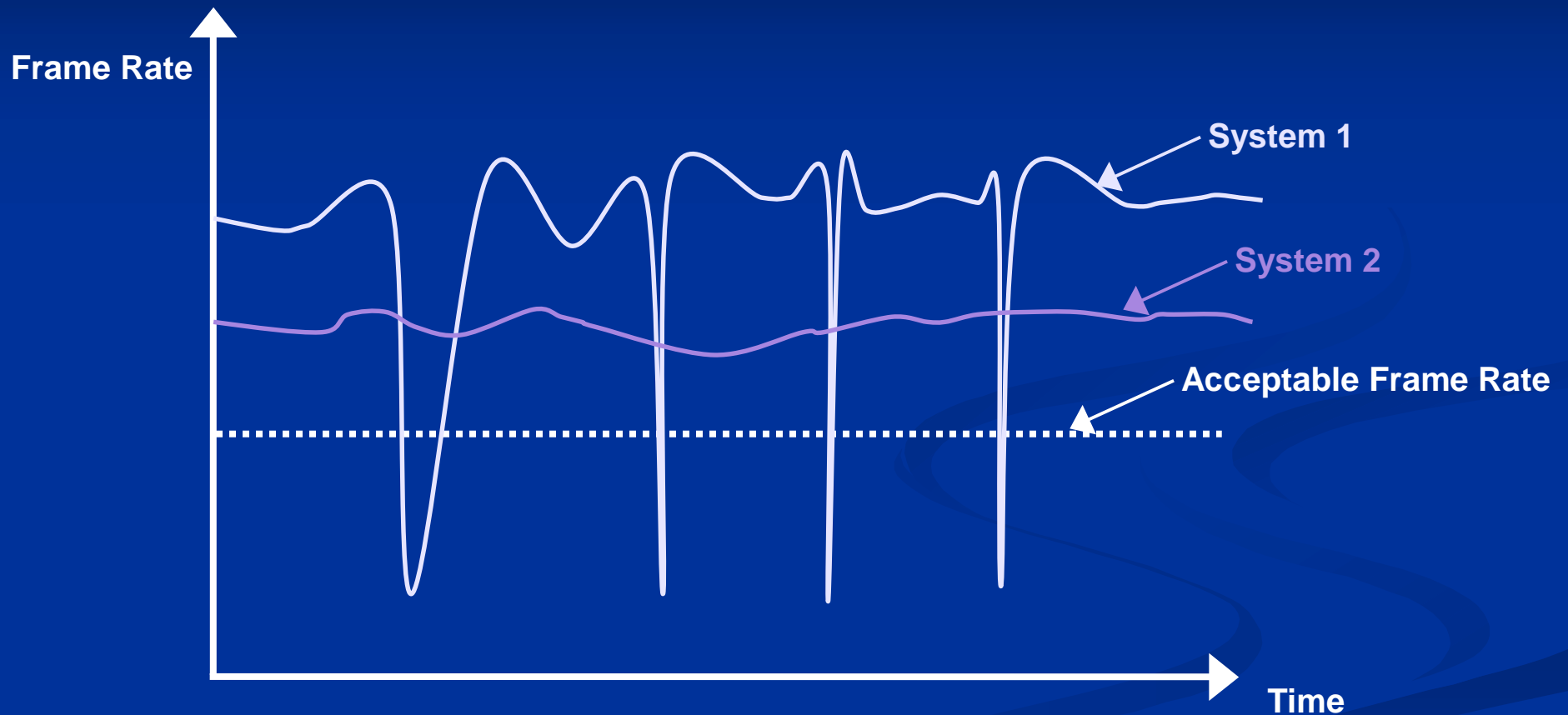
- Environment (changing usage, pervasive computing, portability/interoperability, cost, relevant to who?)
- Complexity (more usage models, defining what is being tested, licensing)
- Hardware improvements continue at a tremendous pace (impacts scaling); software tends to lag
- Comprehensive AND Fair Run Rules (especially in light of new technology, enforcement, disclosure)
- Education (one number vs. many)

*An often overlooked fact is that benchmarks need to evolve with time... Or yesterday's good benchmark may NOT be today's good benchmark.
Benchmarking is getting harder...*

Or in other words: How do end-user look at things?

Example: PC Game Testing

Which system is better? (Could also apply to application response time)



Benchmarking needs to continue to evolve as does the benchmarker's awareness; new technologies and usages will require new ways of accounting for performance.

Backup

Level Setting

- Per www.merriamwebster.com: Benchmark = “a standardized problem or test that serves as a basis for evaluation or comparison (as of computer system performance)”
- Examples Of Benchmarks:
 - Car Efficiency: miles per gallon
 - Sports Statistics: ERA, batting average
 - School: GPA, SAT, GRE
- Benchmarks lie between the extremes of “wild guess” and “actual environment”
- Ideally, measure exactly what you want to evaluate but there are issues with: time, money, available data, economies of scale



*“Benchmarks provide successive approximations to reality” – John Mashey
This requires understanding both the benchmark AND your needs!*