Benchmarking Crimes in Systems Security

Erik van der Kouwe, Gernot Heiser, Dennis Andriesse, Herbert Bos, and Cristiano Giuffrida

AMSec fall 2019, 9 Oct 2019, Amsterdam
LATER WÖRK

1

2

3

4
BENCHMARKING FLAWS IN SYSTEMS SECURITY

1. Complete
2. Relevant
3. Sound
4. Reproducible
Selective benchmarking

Improper handling of results

Wrong benchmarks

Benchmarking Flaws in Systems Security
Improper comparison

Benchmarking omissions

Missing information
Benchmarking Flaws in Systems Security
RESYLT

Number of papers

Number of flaws (including underspecified)
KOMMON FLÅWS

- No indication of significance - B4
- Benchmark subsetting - A2
- Only measure run-time overhead - E2
- Throughput degr. ≠ overhead - B2
- No proper baseline - D1
- Not eval. potential perf. degr. - A1
- Incorrect averaging - B5
- Missing software versions - F2
- Missing platform specification - F1
- Bad math - B3
- Inappropriate/misl. benchm. - C2
DISKUSSION

2 ×

49 ×
KVESTIONS?

Acknowledgement: human figures from https://idea-instructions.com/ (CC by-nc-sa 4.0)