

9. REFERENCES

- [1] A. R. Alameldeen, C. J. Mauer, M. Xu, P. J. Harper, M. M. K. Martin, D. J. Sorin, M. D. Hill, and D. A. Wood. Evaluating non-deterministic multi-threaded commercial workloads. In *Proceedings of the Fifth Workshop on Computer Architecture Evaluation Using Commercial Workloads*, pages 30–38, 2002.
- [2] B. Atikoglu, Y. Xu, E. Frachtenberg, S. Jiang, and M. Paleczny. Workload analysis of a large-scale key-value store. In *Proceedings of the 12th Joint Conference On Measurement And Modeling of Computer Systems (SIGMETRICS/Performance'12)*, London, UK, June 2012.
- [3] R. H. Baayen, D. J. Davidson, and D. M. Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412, 2008.
- [4] E. Bakshy and D. Eckles. Uncertainty in online experiments with dependent data: An evaluation of bootstrap methods. In *Proceedings of the 19th ACM SIGKDD conference on knowledge discovery and data mining*. ACM, 2013.
- [5] S. Balakrishnan, R. Rajwar, M. Upton, and K. Lai. The impact of performance asymmetry in emerging multicore architectures. In *Proceedings of the 32nd annual international symposium on Computer Architecture, ISCA'05*, pages 506–517, Washington, DC, USA, 2005. IEEE Computer Society.
- [6] P. Barford and M. Crovella. Generating representative web workloads for network and server performance evaluation. In *Proceedings of the 1998 ACM SIGMETRICS joint International Conference on Measurement and modeling of Computer Systems, SIGMETRICS '98/PERFORMANCE '98*, pages 151–160, New York, NY, USA, 1998. ACM.
- [7] G. E. Box, J. S. Hunter, and W. G. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery*, volume 13. Wiley Online Library, 2005.
- [8] A. Buble, L. Bulej, and P. Tuma. Corba benchmarking: A course with hidden obstacles. In *Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS'03)*, pages 6 pp.–, 2003.
- [9] Y. chung Cheng, U. Hölzle, N. Cardwell, S. Savage, and G. M. Voelker. Monkey see, monkey do: A tool for tcp tracing and replaying. In *USENIX Annual Technical Conference*, pages 87–98, 2004.
- [10] W. G. Cochran. *Sampling techniques*. John Wiley & Sons, 2007.
- [11] L. Eeckhout, H. Vandierendonck, and K. De Bosschere. Designing computer architecture research workloads. *Computer*, 36(2):65–71, 2003.
- [12] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.
- [13] D. G. Feitelson. Workload modeling for computer systems performance evaluation. Unpublished manuscript, v. 0.42. www.cs.huji.ac.il/~feit/wlmod/wlmod.pdf.
- [14] D. G. Feitelson, E. Frachtenberg, and K. L. Beck. Development and Deployment at Facebook. *IEEE Internet Computing*, 17(4), July 2013.
- [15] D. Gupta, K. V. Vishwanath, M. McNett, A. Vahdat, K. Yocum, A. Snoeren, and G. M. Voelker. Diecast: Testing distributed systems with an accurate scale model. *ACM Trans. Comput. Syst.*, 29(2):4:1–4:48, May 2011.
- [16] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for experimental design, measurement, simulation, and modeling*. Wiley, 1991.
- [17] T. Kalibera and R. Jones. Rigorous benchmarking in reasonable time. In *Proceedings of the 2013 International Symposium on International Symposium on Memory Management, ISMM'13*, pages 63–74, New York, NY, USA, 2013. ACM.
- [18] T. Kalibera and P. Tuma. Precise regression benchmarking with random effects: Improving mono benchmark results. In *Formal Methods and Stochastic Models for Performance Evaluation*, volume 4054 of *Lecture Notes in Computer Science*, pages 63–77. Springer Berlin Heidelberg, 2006.
- [19] B. C. Lee and D. M. Brooks. Accurate and efficient regression modeling for microarchitectural performance and power prediction. In *Proceedings of the 12th international conference on Architectural support for programming languages and operating systems, ASPLOS XII*, pages 185–194, New York, NY, USA, 2006. ACM.
- [20] S. Manley, M. Seltzer, and M. Courage. A self-scaling and self-configuring benchmark for web servers (extended abstract). In *Proceedings of the 1998 ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS'98/PERFORMANCE'98*, pages 270–291, New York, NY, USA, 1998. ACM.
- [21] P. McCullagh. Resampling and exchangeable arrays. *Bernoulli*, 6(2):285–301, 2000.
- [22] A. B. Owen. The pigeonhole bootstrap. *The Annals of Applied Statistics*, 1(2):386–411, 2007.
- [23] A. B. Owen and D. Eckles. Bootstrapping data arrays of arbitrary order. *The Annals of Applied Statistics*, 6(3):895–927, 2012.
- [24] D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.
- [25] D. B. Rubin. The Bayesian bootstrap. *The Annals of Statistics*, 9(1):130–134, 1981.
- [26] S. R. Searle, G. Casella, C. E. McCulloch, et al. *Variance Components*. Wiley New York, 1992.
- [27] M. Seltzer, D. Krinsky, K. Smith, and X. Zhang. The case for application-specific benchmarking. In *Workshop on Hot Topics in Operating Systems (HOTOS'99)*, pages 102–107, 1999.
- [28] T. A. Snijders. *Multilevel analysis*. Springer, 2011.