

Reading list for Advanced Operating Systems (Fall 2019)

■ Introduction

- Fred Douglass, M. Frans Kaashoek, John K. Ousterhout, Andrew S. Tanenbaum, Comparison of Two Distributed Systems Amoeba and Sprite. *ACM Transaction on Computer Systems*, Vol.4, 1991.
- LA Barroso, J Dean, U Holzle, The Google cluster architecture *IEEE micro*, 2003.
- Anil Madhavapeddy, Richard Mortier, Charalampos Rotsos, David Scott, Balraj Singh, Thomas Gazagnaire, Steven Smith, Steven Hand, and Jon Crowcroft. 2013. Unikernels: library operating systems for the cloud. *SIGARCH Comput. Archit. News* 41, 1 (March 2013), 461-472.
- Anil Madhavapeddy, Richard Mortier, Charalampos Rotsos, David Scott, Balraj Singh, Thomas Gazagnaire, Steven Smith, Steven Hand, and Jon Crowcroft. 2013. Unikernels: library operating systems for the cloud. *SIGARCH Comput. Archit. News* 41, 1 (March 2013), 461-472.
- Jiamang Wang, Yongjun Wu, Hua Cai, Zhipeng Tang, Zhiqiang Lv, Bin Lu, Yangyu Tao, Chao Li, Jingren Zhou, and Hong Tang, FuxiSort, <http://sortbenchmark.org/FuxiSort2015.pdf>

■ Communication: Message Passing and Remote Procedure Calls

- Bershad, B. N., Anderson, T. E., Lazowska, E. D., and Levy, H. M. 1990. Lightweight remote procedure call. *ACM Trans. Comput. Syst.* 8, 1 (Feb. 1990), 37-55.
- Clark, D. D. and Tennenhouse, D. L. 1990. Architectural considerations for a new generation of protocols. *SIGCOMM Comput. Commun. Rev.* 20, 4 (Aug. 1990), 200-208.
- von Eicken, T., Basu, A., Buch, V., and Vogels, W. 1995. U-Net: a user-level network interface for parallel and distributed computing (includes URL). *SIGOPS Oper. Syst. Rev.* 29, 5 (Dec. 1995), 40-53.
- von Eicken, T., Culler, D.E., Goldstein, S.C., and Schauser, K.E. Active Messages: A Mechanism for Integrated Communication and Computation, in the Proceedings of The 19th Annual International Symposium on Computer Architecture, 1992. Proceedings, page(s): 256-266, 1992.
- V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in Proceedings of the 5th international conference on Emerging networking experiments and technologies - CoNEXT '09, 2009.
- Lixia Zhang, Alexander Afanasyev, Jeffrey Burke, Van Jacobson, kc claffy, Patrick Crowley, Christos Papadopoulos, Lan Wang, and Beichuan Zhang. 2014. Named data networking. *SIGCOMM Comput. Commun. Rev.* 44, 3 (July 2014), 66-73.
- Hunt, Tyler, et al. "Ryoan: a distributed sandbox for untrusted computation on secret data." 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). USENIX Association, 2016.
- Chen, Ang, et al. "Detecting Covert Timing Channels with Time-Deterministic Replay." OSDI. 2014.
- Cedric Renggli, Saleh Ashkboos, Mehdi Aghagolzadeh, Dan Alistarh, and Torsten Hoeftler. 2019. SparCML: high-performance sparse communication for machine learning. In Proceedings of the International Conference for High

Performance Computing, Networking, Storage and Analysis (SC '19).
Association for Computing Machinery, New York, NY, USA, Article 11, 1–15.
DOI:<https://doi.org/10.1145/3295500.3356222>

■ Distributed Shared Memory

- Saha, B., Adl-Tabatabai, A., Hudson, R. L., Minh, C., and Hertzberg, B. 2006. McRT-STM: a high performance software transactional memory system for a multi-core runtime. In Proceedings of the *Eleventh ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming* (New York, New York, USA, March 29 - 31, 2006). PPOPP '06. ACM, New York, NY, 187-197.
- Stumm, M. and Zhou, S. 1990. Algorithms Implementing Distributed Shared Memory. *Computer* 23, 5 (May. 1990), 54-64.
- Nitzberg, B. and Lo, V., "Distributed shared memory: a survey of issues and algorithms," *Computer* , vol.24, no.8, pp.52-60, Aug 1991
- Carriero, N. and Gelernter, D. 1989. Linda in context. *Commun. ACM* 32, 4 (Apr. 1989), 444-458.
- Bolosky, W., Fitzgerald, R., and Scott, M. 1989. Simple but effective techniques for NUMA memory management. In Proceedings of the Twelfth ACM Symposium on Operating Systems Principles SOSP '89. ACM, New York, NY, 19-31.
- A. M. Caulfield, L. M. Grupp, and S. Swanson, "Gordon : using flash memory to build fast, power-efficient clusters for data-intensive applications," in *ACM SIGARCH Computer Architecture News*, 2009, vol. 37, no. 1, p. 217.
- Zaharia, Matei, et al. "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing." Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. USENIX Association, 2012.
- J. Fresno, D. Barba, A. Gonzalez-Escribano, and D. R. Llanos, "HitFlow: A Dataflow Programming Model for Hybrid Distributed- and Shared-Memory Systems," *Int. J. Parallel Program.*, 2018.
- Németh, Gábor, Dániel Géhberger, and Péter Mátray. "DAL: A Locality-Optimizing Distributed Shared Memory System." 9th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 17). 2017.
- Kim, B., Heo, S., Lee, G., Park, S., Kim, H., & Kim, J. (2016). Heterogeneous distributed shared memory for lightweight internet of things devices. *IEEE Micro*, 36(6), 16-24.

■ Synchronization

- Shin, K.G. ; Ramanathan, P., Clock synchronization of a large multiprocessor system in the presence of malicious faults, *IEEE Trans. Comput.*; (United States); Journal Volume: C-36:1987.
- Chandy, K. M., Misra, J., and Haas, L. M. 1983. Distributed deadlock detection. *ACM Trans. Comput. Syst.* 1, 2 (May. 1983), 144-156.
- F. Cristian, A probabilistic approach to distributed clock synchronization 9th International Conference on Distributed Computing Systems, 1989. Pages: 288-296.
- Shaoshan Liu, and Jean-Luc Gaudiot, Synchronization Mechanisms on Modern Multi-core Architectures, *Advances in Computer Systems Architecture*, Vol. 4697/2007, Page 290 -- 303.
- V. Jungnickel, T. Wirth, M. Schellmann, T. Haustein, and W. Zirwas, "Synchronization of cooperative base stations," in 2008 IEEE International Symposium on Wireless Communication Systems, 2008, pp. 329–334.

- Corbett, James C., et al. "Spanner: Google's globally distributed database." ACM Transactions on Computer Systems (TOCS) 31.3 (2013): 8.
- Li, Jialin, et al. "Just say NO to Paxos overhead: Replacing consensus with network ordering." 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA. 2016.
- Hoch, Ezra N., et al. "Bizur: A key-value consensus algorithm for scalable file-systems." arXiv preprint arXiv:1702.04242 (2017).
- Wang, W., Zhang, C., Yang, L., Xia, J., Chen, K., & Tan, K. (2020). Divide-and-Shuffle Synchronization for Distributed Machine Learning. arXiv preprint arXiv:2007.03298.
- Distributed Resource Management
 - Process and Thread Management:
 - Thomas E. Anderson, Edward D. Lazowska, and Henry M. Levy, The Performance Implications of Thread Management Alternatives for Shared-Memory Multiprocessors, IEEE Transactions on Computers, Vol. 38, No. 12, Pages 1631-1644, December 1989
 - M. Maheswaran, S. Ali, H. J. Siegel, D. Hensgen, R. F. Freund, "Dynamic mapping of a class of independent tasks onto heterogeneous computing systems," Journal of Parallel and Distributed Computing, Vol. 59, No. 2, Nov 1999, pp. 107-131.
 - R. Raman, M. Livny, M. Solomon, "Matchmaking: Distributed Resource Management for High Throughput Computing," hpdc, p. 140, Seventh IEEE International Symposium on High Performance Distributed Computing (HPDC-7 '98), 1998
 - Russ, S.H. Reece, K. Robinson, J. Meyers, B. Rajan, D. Rajagopalan, L. Chun-Heong Tan, Hector: an agent based architecture for dynamic resource management, in the Proceedings of in: Concurrency, IEEE, Publication Date: Apr-Jun 1999 Volume: 7, Issue: 2, On page(s): 47-55
 - Scheduling:
 - D. L. Black, Scheduling Support for Concurrency and Parallelism in the Mach Operating System, IEEE Computer, 23, 5, Pages 35-43, May 1990.
 - Boutin, Eric, et al. "Apollo: Scalable and Coordinated Scheduling for Cloud-Scale Computing." OSDI. Vol. 14. 2014.
 - Venkataraman, Shivaram, et al. "The Power of Choice in Data-Aware Cluster Scheduling." OSDI. 2014.
 - Q. Fang, J. Wang, Q. Gong and M. Song, "Thermal-Aware Energy Management of an HPC Data Center via Two-Time-Scale Control," in IEEE Transactions on Industrial Informatics, vol. 13, no. 5, pp. 2260-2269, Oct. 2017.
 - Process Migration:
 - M.Theimer, K.Lantz, D.Cherton, "Preemptable Remote Execution", Proceedings of the 10th SOSP, Operating Systems Review, Vol. 19, No. 5, Pages 2-12, December 1985.
 - S.-H. Hung, C.-S. Shih, J.-P. Shieh, C.-P. Lee, and Y.-H. Huang, "An Online Migration Environment for Executing Mobile Applications on the Cloud," in Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on, 2011, pp. 20 - 27.

- Byung-Gon Chun and Petros Maniatis. 2009. Augmented smartphone applications through clone cloud execution. In Proceedings of the 12th conference on Hot topics in operating systems (HotOS'09). USENIX Association, Berkeley, CA, USA, 8-8.
- B.-G. Chun, S. Ihm, P. Maniatis, M. Naik, and A. Patti, "CloneCloud," in Proceedings of the sixth conference on Computer systems - EuroSys '11, 2011, p. 301.
- Umesh Deshpande, Kate Keahey, Traffic-sensitive Live Migration of Virtual Machines, Future Generation Computer Systems, Volume 72, 2017, Pages 118-128.
- Distributed Computing:
 - Jim Basney and Miron Livny, "Deploying a High Throughput Computing Cluster", High Performance Cluster Computing, Rajkumar Buyya, Editor, Vol. 1, Chapter 5, Prentice Hall PTR, May 1999. More information about Condor is available at its homepage <http://www.cs.wisc.edu/condor/>
- Avasalcai, C., & Dustdar, S. (2019, March). Latency-aware distributed resource provisioning for deploying iot applications at the edge of the network. In Future of Information and Communication Conference (pp. 377-391). Springer, Cham.
- Lahbib, A., Toumi, K., Laouiti, A., & Martin, S. (2019, September). DRMF: a Distributed Resource Management Framework for industry 4.0 environments. In 2019 IEEE 18th International Symposium on Network Computing and Applications (NCA) (pp. 1-9). IEEE.
- Distributed File Systems
 - Athicha Muthitacharoen, Robert Morris, Thomer M. Gil, Bengie Chen, Ivy: A Read/Write Peer-to-Peer File System (2002)
 - NFS: NFS Version 4 Papers are here: <http://www.nfsv4.org/>. The NFS v4 RFC is <http://www.ietf.org/rfc/rfc3010.txt> here.
 - R. Sandberg, D. Goldberg, S. Kleiman, D. Walsh, and B. Lyon. Design and Implementation of the Sun Network Filesystem. Proc. USENIX Summer Conference, June 1985, pp. 119-130.
 - M. N. Nelson, B. B. Welch, and J. K. Ousterhout. Caching in the Sprite Network File System, ACM Transactions on Computer Systems, 6(1), February 1988, pp. 134-154.
 - J. H. Howard, M. L. Kazar, S. G. Menees, D. A. Nichols, M. Satyanarayanan, R. N. Sidebotham, and M. J. West. Scale and Performance in a Distributed File System. *ACM Transactions on Computer Systems*, 6(1), February 1988, pp. 51-81
 - T. E. Anderson, M. D. Dahlin, J. M. Neefe, D. A. Patterson, D. S. Roselli, and R. Y. Wang. Serverless Network File Systems. *Proc. 15th SOSP, ACM Transactions on Computer Systems*, 14(1), February 1996, pp. 41-79.
 - Chandramohan Thekkath, Timothy Mann, and Edward Lee. Frangipani: A Scalable Distributed File System. Proc. of the 16th ACM Symposium on Operating Systems Principles, October 1997, pages 224-237.
 - G. Kuenning and G. Popek. Automated Hoarding for Mobile Computers. Proc. of the 16th ACM Symposium on Operating Systems Principles, October 1997
 - D. B. Terry, M. M. Theimer, K. Petersen, A. J. Demers, M. J. Spreitzer, and C. Hauser. Managing Update Conflicts in Bayou, a Weakly Connected Replicated Storage System *Proceedings 15th Symposium on Operating Systems Principles (SOSP-15)*, Copper Mountain, Colorado, December 1995, pages 172-183.

- Bruce Walker, Gerald Popek, Robert English, Charles Kline, Greg Thiel, The LOCUS Distributed Operating System, *9th Symposium on Operating Systems Principles (SOSP)*, Bretton Woods, New Hampshire, November 1983, pp. 49-70.
- M. Satyanarayanan, John H. Howard, David A. Nichols, Robert N. Sidebotham, Alfred Z. Spector, Michael J. West, The ITC Distributed File System: Principles and Design, *10th Symposium on Operating Systems Principles (SOSP)*, Orcas Island, Washington, pp. 35-50. December 1985.
- James J. Kistler, M. Satyanarayanan, Disconnected Operation in the Coda File System, *13th Symposium on Operating Systems Principles*, Asilomar, California, pp. 213-225. October 1991.
- Zebra Network File System
- K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," in 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), 2010, pp. 1–10.
- S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google file system," *ACM SIGOPS Oper. Syst. Rev.*, vol. 37, no. 5, p. 29, Dec. 2003.
- Ardekani, Masoud Saeida, and Douglas B. Terry. "A Self-Configurable Geo-Replicated Cloud Storage System." *OSDI*. 2014.
- Lu, Y., Shu, J., Chen, Y., & Li, T. (2017, July). Octopus: an RDMA-enabled distributed persistent memory file system. In 2017 USENIX Annual Technical Conference (USENIX ATC 17)(pp. 773-785).
- Yang, J., Izraelevitz, J., & Swanson, S. (2019). Orion: A distributed file system for non-volatile main memory and RDMA-capable networks. In 17th {USENIX} Conference on File and Storage Technologies ({FAST} 19) (pp. 221-234).
- Middleware for IoT
 - N. Reijers, Y.-C. Wang, C.-S. Shih, J. Y. Hsu, and K.-J. Lin, "Building intelligent middleware for large scale CPS systems," in 2011 IEEE International Conference on Service-Oriented Computing and Applications (SOCA), 2011, pp. 1–4.
 - Y. Zhang, C. Gill, and C. Lu, "Reconfigurable Real-Time Middleware for Distributed Cyber-Physical Systems with Aperiodic Events," in 2008 The 28th International Conference on Distributed Computing Systems, 2008, pp. 581–588.
 - M. C. Huebscher and J. A. McCann, "An adaptive middleware framework for context-aware applications," *Pers. Ubiquitous Comput.*, vol. 10, no. 1, pp. 12–20, Aug. 2005.
 - Xie, C., Yu, B., Zeng, Z., Yang, Y., & Liu, Q. (2020). Multi-Layer Internet of Things Middleware based on Knowledge Graph. *IEEE Internet of Things Journal*.
- Embedded Real-Time Operating Systems
 - J. Eidson, E. Lee, and S. Matic, "Distributed Real-Time Software for Cyber – Physical Systems," *Proc. IEEE*, vol. 100, no. 1, pp. 45–59, 2012.
 - G. Heiser, "Hypervisors for Consumer Electronics," in 2009 6th IEEE Consumer Communications and Networking Conference, 2009, pp. 1–5.
 - G. Heiser and B. Leslie, "The OKL4 Microvisor: Convergence point of microkernels and hypervisors," in *Proceedings of the first ACM asia-pacific workshop ...*, 2010, pp. 19–23.
 - G. Heiser, "Virtualizing embedded systems: why bother?," in *Proceedings of the 48th Design Automation ...*, 2011, pp. 901–905.
- High Performance Computing:

- Albert Reuther, Chansup Byun, William Arcand, David Bestor, Bill Bergeron, Matthew Hubbell, Michael Jones, Peter Michaleas, Andrew Prout, Antonio Rosa, Jeremy Kepner, Scalable system scheduling for HPC and big data, Journal of Parallel and Distributed Computing, Volume 111, 2018, Pages 76-92, ISSN 0743-7315.
- M. Jones et al., "Performance measurements of supercomputing and cloud storage solutions," 2017 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, 2017, pp. 1-5.
- T. Kanewala, M. Zalewski and A. Lumsdaine, "Distributed-memory fast maximal independent set," 2017 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, 2017, pp. 1-7.
- Robotic Operating Systems:
 - Quigley, Morgan, et al. "ROS: an open-source Robot Operating System." ICRA workshop on open source software. Vol. 3. No. 3.2. 2009.
 - Elkady, Ayssam, and Tarek Sobh. "Robotics middleware: A comprehensive literature survey and attribute-based bibliography." Journal of Robotics 2012 (2012).
 - Ando, Noriaki, et al. "RT-middleware: distributed component middleware for RT (robot technology)." 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2005.