Advanced Operating Systems: List of Papers

Ravi Prakash, University of Texas at Dallas

Spring 2021

Theoretical Foundations

Scalar clock paper by Lamport [20], vector clock paper by Mattern [22], causal ordering paper by Raynal and Schiper [26], the distributed snapshot paper by Chandy and Lamport [5].

Physical Clock Synchronization

Papers by Cristian [7], Gusella and Zatti [16] and Srikanth and Toueg [31]

Mutual Exclusion

Algorithms by Lamport [20], Ricart and Agrawala [28] with the optimization by Carvalho and Roucairol [3], Maekawa [21], and Raymond [25].

Crash and Recovery

Paper by Skeen and Stonebraker discussing crash failures and two-phase commit protocol and its extensions [30], Koo and Toueg [19].

Agreement

The survey paper by Michael Fischer [11] and the Practical Byzantine Fault Tolerance paper by Castro and Liskov [4].

Consistency and Voting

Papers by Davidson et al. [8], Gifford [14] and Jajodia and Mutchler [18].

Managing Large Distributed Data Clusters

Papers on the Google file system [13], Dynamo from Amazon [10].

CAP Theorem

PODC 2002 paper by Gilbert and Lynch [15].

Distributed Shared Memory

The ISCA '94 by Reinhardt et al. [27].

Distributed Scheduling

The SOSP 2009 paper by Isard et al. [17].

Software Transactional Memory

Paper by Shavit and Touitou [29].

Additional Papers (as time permits)

MapReduce [9], Bigtable [6], failure trends in large disk drive population [24], patch-based file system [12], Sinfonia from HP Labs and Vmware [1], VirtualPower [23], Fast Array of Wimpy Nodes [2], blockchain related survey by Vukoli'c [32].

References

[1] M.K. Aguilera, A. Merchant, M. Shah, A. Veitch, and C. Karamanolis. Sinfonia: a new paradigm for bulding scalable distributed systems. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2007.

[2] David G. Andersen, Jason Franklin, Michael Kaminsky, Amar Phanishayee, Lawrence Tan, and Vijay Vasudevan. FAWN: a fast array of wimpy nodes. In *Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles*, SOSP '09, pages 1–14, New York, NY, USA, 2009. ACM.

[3] O.S.F. Carvalho and G. Roucairol. On Mutual Exclusion in Computer Networks (Technical Correspondence). Communications of the ACM, February 1983.

[4] Miguel Castro and Barbara Liskov. Practical byzantine fault tolerance and proactive recovery. *ACM Transactions on Computer Systems (TOCS)*, 20(4):398–461, 2002.

[5] K. M. Chandy and L. Lamport. Distributed Snapshots : Determining Global States of Distributed Systems. *ACM Transactions on Computer Systems*, 3(1):63–75, February 1985.

[6] F. Chang, J. Dean, S. Ghemawat, W.C. Hsieh, D.A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R.E. Gruber. Bigtable: A Distributed Storage System for Structured Data. In *Proceedings of the Seventh Symposium on Operating Systems Design and Implementation (OSDI'06)*. USENIX Association, 2006.

[7] F. Cristian. Probabilistic Clock Synchronization. *Distributed Computing*, 3:146–158, 1989.

[8] S.B. Davidson, H. Garcia-Molina, and D. Skeen. Consistency in Partitioned Networks. *ACM Computing Surveys*, 17(3):341–370, September 1985.

[9] J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. In *Proceedings of the Sixth Symposium on Operating Systems Design and Implementation (OSDI'04)*, pages 137–149. USENIX Association, 2004.

[10] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels. Dynamo: Amazon's Highly Available Key-value Store. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2007.

[11] Michael Fischer. The consensus problem in unreliable distributed systems (a brief survey). In Marek Karpinski, editor, *Foundations of Computation Theory*, volume 158 of *Lecture Notes in Computer Science*, pages 127–140. Springer Berlin / Heidelberg, 1983.

[12] C. Frost, M. Mammarella, E. Kohler, A. de los Reyes, S. Hovsepian, A. Matsuoka, and L. Zhang. Generalized File System Dependencies. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2007.

[13] S. Ghemawat, H. Gobioff, and S.-T. Leung. The Google File System. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2003.

[14] D.K. Gifford. Weighted Voting for Replicated Data. In *Proceedings of the 7th Symposium on Operating Systems Principles*, pages 150–162. ACM, 1979.

[15] Seth Gilbert and Nancy Lynch. Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. *Acm Sigact News*, 33(2):51–59, 2002.

[16] R. Gusella and S. Zatti. The Accuracy of the Clock Synchronization Achieved by TEMPO in Berkeley UNIX 4.3BSD. *IEEE Transactions on Software Engineering*, 15(7):847–853, July 1989.

[17] Michael Isard, Vijayan Prabhakaran, Jon Currey, Udi Wieder, Kunal Talwar, and Andrew Goldberg. Quincy: fair scheduling for distributed computing clusters. In *Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles*, SOSP '09, pages 261–276, New York, NY, USA, 2009. ACM.

[18] S. Jajodia and D. Mutchler. A Hybrid Replica Control Algorithm Combining Static and Dynamic Voting. *IEEE Transactions on Knowledge and Data Engineering*, 1(4):459–469, December 1989.

[19] R. Koo and S. Toueg. Checkpointing and Rollback-Recovery for Distributed Systems. *IEEE Transactions on Software Engineering*, SE-13(1):23–31, January 1987.

[20] L. Lamport. Time, Clocks and the Ordering of Events in a Distributed System. *Communications of the ACM*, 21(7):558–565, July 1978.

[21] M. Maekawa. A $\sqrt{N}$ Algorithm for Mutual Exclusion in Decentralized Systems. *ACM Transactions on Computer Systems*, pages 145–159, May 1985.

[22] F. Mattern. Virtual Time and Global States of Distributed Systems. In M.Cosnard et. al., editor, *Proceedings of the Workshop on Parallel and Distributed Algorithm*, pages 215–226. Elsevier Science Publishers B.V.(North-Holland), 1989.

[23] R. Nathuji and K. Schwan. VirtualPower: Coordinated Power Management in Virtualized Enterprise Systems. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2007.

[24] E. Pinheiro, W.-D. Weber, and L.A. Barroso. Failure Trends in a Large Disk Drive Population. In *Proceedings of Fifth USENIX Conference on File and Storage Technologies (FAST'07)*, February 2007.

[25] K. Raymond. A Tree-Based Algorithm for Distributed Mutual Exclusion. *ACM Transactions on Computer Systems*, 7:61–77, February 1989.

[26] M. Raynal and A. Schiper. The causal ordering abstraction and a simple way to implement it. Technical Report 1132, INRIA, http://hal.archives-ouvertes.fr/docs/00/07/54/27/PDF/RR-1132.pdf, December 1989.

[27] S.K. Reinhardt, J.R. Larus, and D.A. Wood. Tempest and typhoon: user-level shared memory. In *Proceedings of the 21st annual international symposium on Computer architecture*, ISCA '94, pages 325–336, Los Alamitos, CA, USA, 1994. IEEE Computer Society Press.

[28] G. Ricart and A. K. Agrawala. An Optimal Algorithm for Mutual Exclusion in Computer Networks. *Communications of the ACM*, 24(1):9–17, January 1981.

[29] Nir Shavit and Dan Touitou. Software transactional memory. *Distributed Computing*, 10(2):99–116, 1997.

[30] D. Skeen and M. Stonebraker. A Formal Model of Crash Recovery in a Distributed Systems. *IEEE Transactions on Software Engineering*, SE-9(3):219–228, May 1983.

[31] T.K. Srikanth and S. Toueg. Optimal Clock Synchronization. *Journal of the ACM*, January 1987.

[32] Marko Vukolić. The quest for scalable blockchain fabric: Proof-of-work vs. bft replication. In *International Workshop on Open Problems in Network Security*, pages 112–125. Springer, 2015.