

分布式算法课程延伸阅读¹

黄 宇

<http://cs.nju.edu.cn/yuhuang>

¹这是分布式算法课程推荐大家延伸阅读的文献。这个文献列表会在平时不断更新调整，包括增删条目、修改注释、调整文档的目录结构等。这些文献不是一个充分的覆盖，更像是一些经典文献的例举。最新版本请参考“<https://github.com/alg-nju/intro-disalg-course>”。

目录

第一章 建模和模型	2
第二章 Impossibility Results	3
第三章 MSG算法	4
第四章 SHM算法	5
第五章 经典系统	6
第六章 形式化规约与验证	7

第一章 建模和模型

[Lam78] 消息传递模型的基本概念。

[Mat89] 逻辑时钟。

[Lam86a, Lam86b] 基础建模。包括atomic/regular/safe register的概念等。

[Mis86] 硬件访问的公理化建模。

[CT96] Failure detector的概念。

第二章 Impossibility Results

[AE14] 不可能性结果的系统讲解书籍。

[FLP85] FLP结果。

第三章 MSG算法

[Lam01] Paxos算法。

[OO14] Raft算法。

第四章 SHM算法

[ABND95] 定义distributed atomic register。

[HW90] 定义linearizability。

[SWPL11] 多写register的regularity的定义。

[Her91] Consensus number的概念。

[HS11] Progress condition概念的系统阐述。

[Lam74] Bakery算法。

第五章 经典系统

[GL12] CAP 定理，由Eric Brewer在2000年提出，此处是12年后Gilbert和Lynch对CAP定理所作的阐释。

[Aba12] PACELC权衡，可以看做是CAP定理的拓展。

[LH89] 分布共享内存系统。

[CBPS10] 涉及复制技术的经典理论、系统的系统阐述。

[CGR07] Google的Paxos实现。

[CDG⁺08] Google的BigTable。

[BBC⁺11] Google的Megastore。

[CDE⁺12] Google的Spanner。

[DHJ⁺07] Amazon的Dynamo。

[LM10] Facebook的Cassandra。

[HKJR10] Yahoo的Zookeeper。

[Bur06] Google的chubby。

[Gel85] 基于tuple space的协同。

第六章 形式化规约与验证

[Bur14] vis-ar框架: Replicated Data Type的declarative的规约框架。

[CG18] vis-ar框架向分布事务isolation level的拓展。

[MSB17] 共识算法的共性提炼与验证。

参考文献

- [Aba12] Daniel J. Abadi. Consistency tradeoffs in modern distributed database system design: Cap is only part of the story. *Computer*, 45(2):37–42, Feb 2012.
- [ABND95] Hagit Attiya, Amotz Bar-Noy, and Danny Dolev. Sharing memory robustly in message-passing systems. *J. ACM*, 42(1):124–142, January 1995.
- [AE14] Hagit Attiya and Faith Ellen. *Impossibility Results for Distributed Computing*. Morgan & Claypool, 2014.
- [BBC⁺11] Jason Baker, Chris Bond, James C. Corbett, JJ Furman, Andrey Khorlin, James Larson, Jean-Michel Leon, Yawei Li, Alexander Lloyd, and Vadim Yushprakh. Megastore: providing scalable, highly available storage for interactive services. In *Proc. CIDR’11, Conference on Innovative Data System Research*, pages 223–234, 2011.
- [Bur06] Mike Burrows. The Chubby lock service for loosely-coupled distributed systems. In *Proc. OSDI’06, USENIX Symposium on Operating Systems Design and Implementation*, pages 335–350. USENIX, 2006.
- [Bur14] Sebastian Burckhardt. Principles of eventual consistency. *Found. Trends Program. Lang.*, 1(1–2):1–150, October 2014.
- [CBPS10] Bernadette Charron-Bost, Fernando Pedone, and André Schiper, editors. *Replication: Theory and Practice*. Springer-Verlag, Berlin, Heidelberg, 2010.
- [CDE⁺12] James C. Corbett, Jeffrey Dean, Michael Epstein, Andrew Fikes, Christopher Frost, J. J. Furman, Sanjay Ghemawat, Andrey Gubarev, Christopher Heiser, Peter Hochschild, Wilson Hsieh, Sebastian Kanthak, Eugene Kogan, Hongyi Li, Alexander Lloyd, Sergey Melnik, David Mwaura, David Nagle, Sean Quinlan, Rajesh Rao, Lindsay Rolig, Yasushi Saito, Michal Szymaniak, Christopher Taylor, Ruth Wang, and Dale Woodford. Spanner: Google’s globally-distributed database. In *Proc. OSDI’12, USENIX Symposium on Operating Systems Design and Implementation*, pages 251–264. USENIX, 2012.

- [CDG⁺08] Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, and Robert E. Gruber. Bigtable: A distributed storage system for structured data. *ACM Trans. Comput. Syst.*, 26(2):4:1–4:26, June 2008.
- [CG18] Andrea Cerone and Alexey Gotsman. Analysing snapshot isolation. *J. ACM*, 65(2), January 2018.
- [CGR07] Tushar D. Chandra, Robert Griesemer, and Joshua Redstone. Paxos made live: An engineering perspective. In *Proceedings of the Twenty-sixth Annual ACM Symposium on Principles of Distributed Computing*, PODC '07, pages 398–407. ACM, 2007.
- [CT96] Tushar Deepak Chandra and Sam Toueg. Unreliable failure detectors for reliable distributed systems. *J. ACM*, 43(2):225–267, March 1996.
- [DHJ⁺07] Giuseppe DeCandia, Deniz Hastorun, Madan Jampani, Gunavardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Voss, and Werner Vogels. Dynamo: Amazon’s highly available key-value store. In *Proceedings of Twenty-first ACM SIGOPS Symposium on Operating Systems Principles*, SOSP '07, pages 205–220. ACM, 2007.
- [FLP85] Michael J. Fischer, Nancy A. Lynch, and Michael S. Paterson. Impossibility of distributed consensus with one faulty process. *J. ACM*, 32(2):374–382, April 1985.
- [Gel85] David Gelernter. Generative communication in linda. *ACM Trans. Program. Lang. Syst.*, 7(1):80–112, January 1985.
- [GL12] Seth Gilbert and Nancy A. Lynch. Perspectives on the cap theorem. *Computer*, 45(2):30–36, 2012.
- [Her91] Maurice Herlihy. Wait-free synchronization. *ACM Trans. Program. Lang. Syst.*, 13(1):124–149, January 1991.
- [HKJR10] Patrick Hunt, Mahadev Konar, Flavio P. Junqueira, and Benjamin Reed. ZooKeeper: wait-free coordination for internet-scale systems. In *Proc. ATC'10, USENIX Annual Technical Conference*, pages 145–158. USENIX, 2010.
- [HS11] Maurice Herlihy and Nir Shavit. On the nature of progress. In *Proceedings of the 15th International Conference on Principles of Distributed Systems*, OPODIS'11, pages 313–328, 2011.

- [HW90] Maurice P. Herlihy and Jeannette M. Wing. Linearizability: a correctness condition for concurrent objects. *ACM Transactions on Programming Languages and Systems*, 12:463–492, July 1990.
- [Lam74] Leslie Lamport. A new solution of dijkstra’s concurrent programming problem. *Commun. ACM*, 17(8):453–455, August 1974.
- [Lam78] Leslie Lamport. Time, clocks, and the ordering of events in a distributed system. *Commun. ACM*, 21(7):558–565, 1978.
- [Lam86a] Leslie Lamport. On interprocess communication. part i: Basic formalism. *Distributed Computing*, 1(2):77–85, 1986.
- [Lam86b] Leslie Lamport. On interprocess communication. part ii: Algorithms. *Distributed Computing*, 1(2):86–101, 1986.
- [Lam01] Leslie Lamport. Paxos made simple. *ACM SIGACT News (Distributed Computing Column)* 32, 4 (Whole Number 121, December 2001), pages 51–58, December 2001.
- [LH89] Kai Li and Paul Hudak. Memory coherence in shared virtual memory systems. *ACM Trans. Comput. Syst.*, 7(4):321–359, November 1989.
- [LM10] Avinash Lakshman and Prashant Malik. Cassandra: A decentralized structured storage system. *SIGOPS Oper. Syst. Rev.*, 44(2):35–40, April 2010.
- [Mat89] Friedemann Mattern. Virtual time and global states of distributed systems. In *Proc. International Workshop on Parallel and Distributed Algorithms*, pages 215–226, Holland, 1989.
- [Mis86] J. Misra. Axioms for memory access in asynchronous hardware systems. *ACM Trans. Program. Lang. Syst.*, 8(1):142–153, January 1986.
- [MSB17] Ognjen Marić, Christoph Sprenger, and David Basin. Cutoff bounds for consensus algorithms. In Rupak Majumdar and Viktor Kunčák, editors, *Computer Aided Verification*, pages 217–237, Cham, 2017. Springer International Publishing.
- [OO14] Diego Ongaro and John Ousterhout. In search of an understandable consensus algorithm. In *Proceedings of the 2014 USENIX Conference on USENIX Annual Technical Conference*, USENIX ATC’14, pages 305–320, Berkeley, CA, USA, 2014. USENIX Association.
- [SWPL11] Cheng Shao, Jennifer L. Welch, Evelyn Pierce, and Hyunyoung Lee. Multiwriter consistency conditions for shared memory registers. *SIAM J. Comput.*, 40(1):28–62, January 2011.