分布式算法延伸阅读 [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

文献 1.1 ([Lamport, 1978])**.** 消息传递模型的基本概念。

文献 1.2 ([Mattern, 1989])**.** 逻辑时钟。

文献 1.3 ([Lamport, 1986a, Lamport, 1986b])**.** 基础建模。
包括atomic/regular/safe register的概念等。

文献 1.4 ([Misra, 1986])**.** 硬件访问的公理化建模。

文献 1.5 ([Chandra and Toueg, 1996])**.** Failure detector的概念。

第二章　Impossibility Results

从道理上讲，“impossibility results”划出了理论的边界，我们才可以进而研究后续的算法设计与分析。这也是为什么impossibility results被放在了具体算法的前面。

但是实际上理论结果的形成，包括大家学习分布式算法的过程，却往往不是这样的。所以大家可以尽管跳过这一章，先了解后面的内容。待对分布式算法有了基本的认识之后，再回来深入研究经典的impossibility results。

文献 2.1 ([Attiya and Ellen, 2014])**.** 不可能性结果的系统讲解书籍。

文献 2.2 ([Fischer et al., 1985])**.** FLP结果。

第三章　MSG算法

文献 3.1 ([Lamport, 2001])**.** Paxos算法。

文献 3.2 ([Ongaro and Ousterhout, 2014])**.** Raft算法。

第四章　SHM算法

文献 4.1 ([Attiya et al., 1995])**.** 定义distributed atomic register。

文献 4.2 ([Herlihy and Wing, 1990])**.** 定义linearizability。

文献 4.3 ([Shao et al., 2011])**.** 多写register的regularity的定义。

文献 4.4 ([Herlihy, 1991])**.** Consensus number的概念。

文献 4.5 ([Herlihy and Shavit, 2011])**.** Progress condition概念的系统阐述。

文献 4.6 ([Lamport, 1974])**.** Bakery算法。

第五章　经典系统

文献 5.1 ([Gilbert and Lynch, 2012, Brewer, 2012])．CAP 定理。
Eric Brewer在2000年提出。
[Gilbert and Lynch, 2012]是12年后Gilbert和Lynch对CAP定理所作的阐释。
[Brewer, 2012]是作者自己在12年后的思考。

文献 5.2 ([Abadi, 2012])．PACELC权衡，可以看作是CAP定理的拓展。

文献 5.3 ([Li and Hudak, 1989])．分布共享内存系统。

文献 5.4 ([Charron-Bost et al., 2010])．涉及复制技术的经典理论、系统的系统阐述。

文献 5.5 ([Chandra et al., 2007])．Google的Paxos实现。

文献 5.6 ([Chang et al., 2008])．Google的BigTable。

文献 5.7 ([Baker et al., 2011])．Google的Megastore。

文献 5.8 ([Corbett et al., 2012])．Google的Spanner。

文献 5.9 ([DeCandia et al., 2007])．Amazon的Dynamo。

文献 5.10 ([Lakshman and Malik, 2010])．Facebook的Cassandra。

文献 5.11 ([Hunt et al., 2010])．Yahoo的Zookeeper。

文献 5.12 ([Burrows, 2006])．Google的chubby。

文献 5.13 ([Gelernter, 1985])．基于tupple space的协同。

第六章　形式化规约与验证

文献 6.1 ([Burckhardt, 2014])**.** vis-ar框架。
Replicated Data Type的declarative的规约框架。

文献 6.2 ([Cerone and Gotsman, 2018])**.** vis-ar框架向分布事务isolation level的拓展。

文献 6.3 ([Crooks et al., 2017])**.** 事务isolation level的规约框架。
基于状态的。数据库领域的做法。与vis-ar框架形成对照。

文献 6.4 ([Marić et al., 2017])**.** 共识算法的共性提炼与验证。

参考文献

[Abadi, 2012] Abadi, D. J. (2012). Consistency tradeoffs in modern distributed database system design: Cap is only part of the story. *Computer*, 45(2):37–42.

[Attiya et al., 1995] Attiya, H., Bar-Noy, A., and Dolev, D. (1995). Sharing memory robustly in message-passing systems. *J. ACM*, 42(1):124–142.

[Attiya and Ellen, 2014] Attiya, H. and Ellen, F. (2014). *Impossibility Results for Distributed Computing*. Morgan & Claypool.

[Baker et al., 2011] Baker, J., Bond, C., Corbett, J. C., Furman, J., Khorlin, A., Larson, J., Leon, J.-M., Li, Y., Lloyd, A., and Yushprakh, V. (2011). Megastore: providing scalable, highly available storage for interactive services. In *Proc. CIDR'11, Conference on Innovative Data System Research*, pages 223–234.

[Brewer, 2012] Brewer, E. (2012). Cap twelve years later: How the "rules" have changed. *Computer*, 45(2):23–29.

[Burckhardt, 2014] Burckhardt, S. (2014). Principles of eventual consistency. *Found. Trends Program. Lang.*, 1(1–2):1–150.

[Burrows, 2006] Burrows, M. (2006). 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.

[Cerone and Gotsman, 2018] Cerone, A. and Gotsman, A. (2018). Analysing snapshot isolation. *J. ACM*, 65(2).

[Chandra et al., 2007] Chandra, T. D., Griesemer, R., and Redstone, J. (2007). 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.

[Chandra and Toueg, 1996] Chandra, T. D. and Toueg, S. (1996). Unreliable failure detectors for reliable distributed systems. *J. ACM*, 43(2):225–267.

[Chang et al., 2008] Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., Chandra, T., Fikes, A., and Gruber, R. E. (2008). Bigtable: A distributed storage system for structured data. *ACM Trans. Comput. Syst.*, 26(2):4:1–4:26.

[Charron-Bost et al., 2010] Charron-Bost, B., Pedone, F., and Schiper, A., editors (2010). *Replication: Theory and Practice*. Springer-Verlag, Berlin, Heidelberg.

[Corbett et al., 2012] Corbett, J. C., Dean, J., Epstein, M., Fikes, A., Frost, C., Furman, J. J., Ghemawat, S., Gubarev, A., Heiser, C., Hochschild, P., Hsieh, W., Kanthak, S., Kogan, E., Li, H., Lloyd, A., Melnik, S., Mwaura, D., Nagle, D., Quinlan, S., Rao, R., Rolig, L., Saito, Y., Szymaniak, M., Taylor, C., Wang, R., and Woodford, D. (2012). Spanner: Google's globally-distributed database. In *Proc. OSDI'12, USENIX Symposium on Operating Systems Design and Implementation*, pages 251–264. USENIX.

[Crooks et al., 2017] Crooks, N., Pu, Y., Alvisi, L., and Clement, A. (2017). Seeing is believing: A client-centric specification of database isolation. In *Proceedings of the ACM Symposium on Principles of Distributed Computing*, PODC'17, page 73–82, New York, NY, USA. Association for Computing Machinery.

[DeCandia et al., 2007] DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Vosshall, P., and Vogels, W. (2007). 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.

[Fischer et al., 1985] Fischer, M. J., Lynch, N. A., and Paterson, M. S. (1985). Impossibility of distributed consensus with one faulty process. *J. ACM*, 32(2):374–382.

[Gelernter, 1985] Gelernter, D. (1985). Generative communication in linda. *ACM Trans. Program. Lang. Syst.*, 7(1):80–112.

[Gilbert and Lynch, 2012] Gilbert, S. and Lynch, N. A. (2012). Perspectives on the cap theorem. *Computer*, 45(2):30–36.

[Herlihy, 1991] Herlihy, M. (1991). Wait-free synchronization. *ACM Trans. Program. Lang. Syst.*, 13(1):124–149.

[Herlihy and Shavit, 2011] Herlihy, M. and Shavit, N. (2011). On the nature of progress. In *Proceedings of the 15th International Conference on Principles of Distributed Systems*, OPODIS'11, pages 313–328.

[Herlihy and Wing, 1990] Herlihy, M. P. and Wing, J. M. (1990). Linearizability: a correctness condition for concurrent objects. *ACM Transactions on Programming Languages and Systems*, 12:463–492.

[Hunt et al., 2010] Hunt, P., Konar, M., Junqueira, F. P., and Reed, B. (2010). ZooKeeper: wait-free coordination for internet-scale systems. In *Proc. ATC'10, USENIX Annual Technical Conference*, pages 145–158. USENIX.

[Lakshman and Malik, 2010] Lakshman, A. and Malik, P. (2010). Cassandra: A decentralized structured storage system. *SIGOPS Oper. Syst. Rev.*, 44(2):35–40.

[Lamport, 1974] Lamport, L. (1974). A new solution of dijkstra's concurrent programming problem. *Commun. ACM*, 17(8):453–455.

[Lamport, 1978] Lamport, L. (1978). Time, clocks, and the ordering of events in a distributed system. *Commun. ACM*, 21(7):558–565.

[Lamport, 1986a] Lamport, L. (1986a). On interprocess communication. part i: Basic formalism. *Distributed Computing*, 1(2):77–85.

[Lamport, 1986b] Lamport, L. (1986b). On interprocess communication. part ii: Algorithms. *Distributed Computing*, 1(2):86–101.

[Lamport, 2001] Lamport, L. (2001). Paxos made simple. *ACM SIGACT News (Distributed Computing Column) 32, 4 (Whole Number 121, December 2001)*, pages 51–58.

[Li and Hudak, 1989] Li, K. and Hudak, P. (1989). Memory coherence in shared virtual memory systems. *ACM Trans. Comput. Syst.*, 7(4):321–359.

[Marić et al., 2017] Marić, O., Sprenger, C., and Basin, D. (2017). Cutoff bounds for consensus algorithms. In Majumdar, R. and Kunčak, V., editors, *Computer Aided Verification*, pages 217–237, Cham. Springer International Publishing.

[Mattern, 1989] Mattern, F. (1989). Virtual time and global states of distributed systems. In *Proc. International Workshop on Parallel and Distributed Algorithms*, pages 215–226, Holland.

[Misra, 1986] Misra, J. (1986). Axioms for memory access in asynchronous hardware systems. *ACM Trans. Program. Lang. Syst.*, 8(1):142–153.

[Ongaro and Ousterhout, 2014] Ongaro, D. and Ousterhout, J. (2014). 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. USENIX Association.

[Shao et al., 2011] Shao, C., Welch, J. L., Pierce, E., and Lee, H. (2011). Multiwriter consistency conditions for shared memory registers. *SIAM J. Comput.*, 40(1):28–62.