分布式算法讲义

黄 宇

2020 年 4 月 10 日

目录

第一部分

计算模型

第一章　分布式算法简介

[3]
多数据中心平台，从硬件设施，到软件基础设施(infrastructure)的介绍。

第二章　分布式计算模型

计算模型的基础是抽象。首先介绍各种抽象。然后介绍由各种不同抽象，组合而来的各种模型 [5]。

第二部分

消息传递算法

第三章　MSG算法1

第四章　Paxos及其变体

// ？ Raft是单独列一张，还是看作Paxos的变体一起讲？
// 有一些变体，例如ParallelRaft，会综合Paxos和Raft的技术

第五章　VR和PBFT

5.1　Viewstamped Replication

// 按照[6, Chap.7]的方式阐述。

5.2　Practical Byzantine Fault Tolerance

PBFT的历史。

首先理解PBFT的一个简化版本。最终的版本有很多实际的细节，与理解算法的精髓是关联不紧密的。

完整的PBFT。

[6, Chap.7]

第三部分

共享存储算法

第四部分

进阶专题

第六章　专题1

【virtual synchrony】

process groups，group membership。

virtual synchrony。

[6, Chap.6]

Process groups are a powerful tool for the developer. They can have names, much like files, and this allows them to be treated like topics in a publish-subscribe system.

One thinks of a process group as a kind of object (abstract data type), and the processes that join the group as importing a replica of that object. Virtual synchrony standardizes the handling of group membership: the system tracks group members, and informs members each time the membership changes, an event called a view change.

【混合BFT】

有些机器只会crash，不会叛变 [10]。

区块链领域也使用这个假设，来提升区块链共识的速度 [8]。

提供满足这种假设的off-the-shelf hardware systems，例如Intel的Software Guard Extensions (SGX) [9]。

第五部分

实际案例

第七章　案例1

【系统类：cloud data store】

[4]

对于cloud data store的介绍。

分布式系统中(主要是cloud data store中)对于ordering of events的tracking。弱一致系统，强一致系统中的clock的设计。

【工具类：TLA+】

[1]。

【工具：benchmarking，评测】

YCSB [7, 2]

素材

参考文献

[1] `https://github.com/tlaplus`.

[2] `https://github.com/brianfrankcooper/YCSB`.

[3] L. A. Barroso, U. Holzle, P. Ranganathan, and M. Martonosi. *The Datacenter As a Computer: Designing Warehouse-Scale Machines*. Morgan & Claypool Publishers, 3rd edition, 2018.

[4] M. Bravo, N. Diegues, J. Zeng, P. Romano, and L. Rodrigues. On the use of clocks to enforce consistency in the cloud. *IEEE Data Eng. Bull.*, 38:18–31, 01 2015.

[5] C. Cachin, R. Guerraoui, and L. Rodrigues. *Introduction to Reliable and Secure Distributed Programming*. Springer Publishing Company, Incorporated, 2nd edition, 2011.

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

[7] B. F. Cooper, A. Silberstein, E. Tam, R. Ramakrishnan, and R. Sears. Benchmarking cloud serving systems with ycsb. In *Proceedings of the 1st ACM Symposium on Cloud Computing*, SoCC '10, pages 143–154, New York, NY, USA, 2010. ACM.

[8] H. Dang, T. T. A. Dinh, D. Loghin, E.-C. Chang, Q. Lin, and B. C. Ooi. Towards scaling blockchain systems via sharding. In *Proceedings of the 2019 International Conference on Management of Data*, SIGMOD ' 19, page 123–140, New York, NY, USA, 2019. Association for Computing Machinery.

[9] F. McKeen, I. Alexandrovich, A. Berenzon, C. V. Rozas, H. Shafi, V. Shanbhogue, and U. R. Savagaonkar. Innovative instructions and software model for isolated execution. In *Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy*, HASP ' 13, New York, NY, USA, 2013. Association for Computing Machinery.

[10] I. Vukotic, V. Rahli, and P. Esteves-Veríssimo. Asphalion: Trustworthy shielding against byzantine faults. *Proc. ACM Program. Lang.*, 3(OOPSLA), Oct. 2019.