

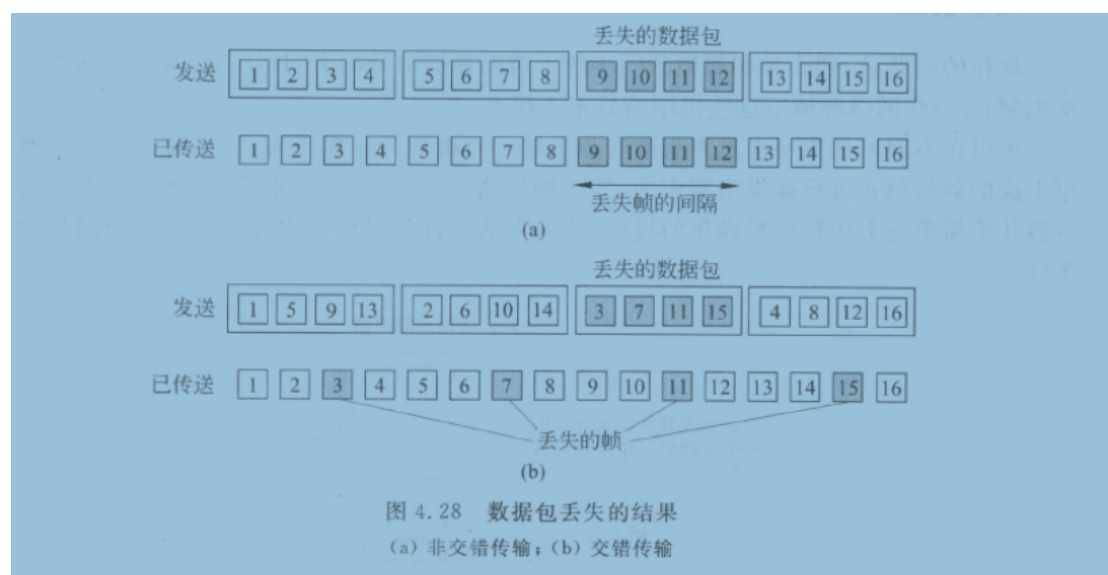
## 2020 年分布式系统考卷回忆

作者：朱泓全，赖伟

- 1、网络操作系统和分布式系统（基于中间件的）的定义分别是什么？请通过一个实例说明这两者之间的关系。
- 2、判断以下三项是属于“策略”还是“机制”。如果是策略，写出相应机制；如果是机制，给出任意一个相应的策略。(类似 2019 年的题目，只要记住机制有哪些就好写)
  - a) 根据运算能力不同将任务分配到不同计算机上
  - b) 支持不同类型的存储方式
  - c) 能够通过多条路径传递数据包
- 3、电子邮件采用什么样的通信模式？简要说明如何在一班 TCP/IP 协议基础上实现这样的通信。
- 4、如何在 TCP 的基础上，实现远程过程调用(RPC)的“最多执行一次”机制。
- 5、客户端没有收到来自服务器的响应，可能的情况有哪些？并给出可能的解决方案。
- 6、设计一个协议让节点的时间是所有节点时间的平均值，详细描述流程以及时间计算方式

第二页：

- 7、为什么 Lamport 不能全序逻辑而向量时钟可以。
- 8、用向量时钟表示下面的传播过程，要求给传播过程标注向量时钟的时间戳，并给出其中的一个错误纠正
- 9、设计一个协议，利用检查点来达成一致性协议。并说明正确的理由
- 10、分别描述流媒体的非交叉传输和交叉传输的优缺点。



- 11、容错问题

(1) 某个没有协调者的非集中式系统中，正确运行的进程必然返回唯一正确结果，进程出错后会返回任意的错误结果，请设计一个协议保证该系统达到  $K$  容错，也就是当  $K$  个进程出错时，进程间仍能达成一致并返回正确结果。(5 分)

需要  $(2K+1)$  个进程

题目要求设计一个协议保证具有拜占庭错误的系统达到  $K$  容错，可以使用  $2K+1$  个重复的进程，当有  $K$  个进程出错时，剩下的  $K+1$  个进程仍然会返回相同的唯一正确结果，只需要取

2K+1 个进程中大多数进程返回的结果就能保证 K 容错。

(2) 有多个传感器从不同位置监控一个系统的温度，该系统的最终温度取这些传感器的平均值。在一个统计周期内，正常工作的传感器对外总是发送其测定的唯一正确温度值；而出现故障的传感器对外发送的温度值是不确定的任意值，但我们无法仅仅通过收到的温度值来判定传感器是否故障。(10 分)

a) 设计一个协议保证正确的传感器间能协调一致，并得到正确的温度平均值。

b) 如果要求在 K 个传感器出错的情况下，系统仍能正常工作，总共需要安放多少个传感器？

答：

第一步：每个传感器向所有的其他传感器发送它测得的温度值

第二步：每个传感器收集其他传感器发送的温度值，组成一个包含自身在内的所有传感器测得的温度值的向量

第三步：每个传感器向其他所有传感器发送向量

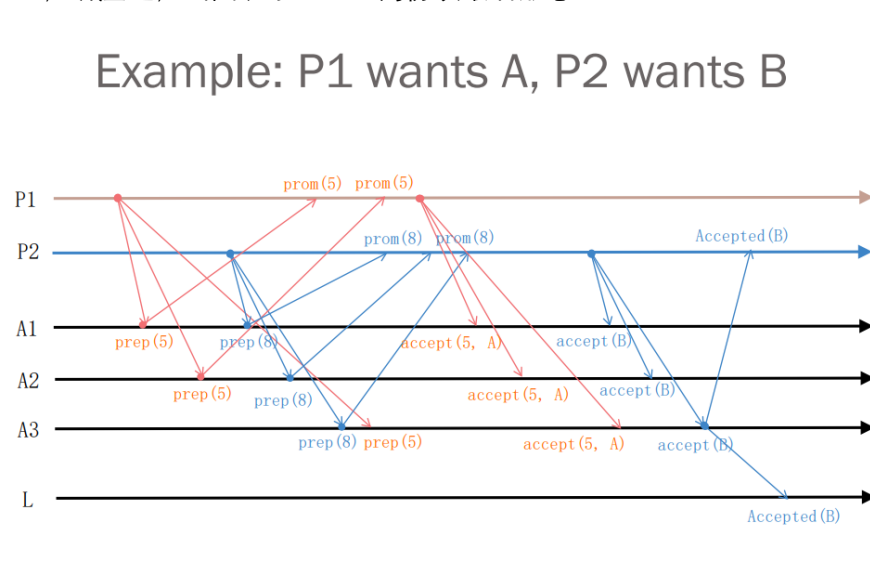
第四步：每个传感器依次检查它收到所有向量的第 i 个元素，如果有某个值 v 占据大多数，那么第 i 个传感器是正常工作的，v 就是它测得的温度值；否则说明第 i 个传感器出现了故障

第五步：每个传感器对所有正常工作的传感器测得的温度值取平均，得到正确的平均温度。

需要(3K+1)个进程

第三页

12, 讲座题, 画图表示 Paxos 的协议最后形态



13, gossip 算法多少论才能终止。可以用复杂度表示，说明理由

14, Hadoop 和 spark 的异同点，为什么 Hadoop Yarn 能提高 Hadoop 的扩放性

15, RFID 中的 Aloha 协议，有哪些时隙，举例说明

16, 类虚拟化和硬件虚拟化的异同之处？虚拟化在云计算中是如何运用的？如何理解基于虚拟化的私有云能提高现有系统性能？