

2019年分布式系统考题

1. 分布式系统的定义与架构（本题满分20分）

(1) 分布式系统的开放性主要体现在哪些方面？（4分）

Be able to interact with services from other open systems, irrespective of the underlying environment:

- Systems should conform to well-defined interfaces
- Systems should support portability of applications
- Systems should easily interoperate
- At least make the distributed system independent from heterogeneity of the underlying environment.

(2) 比特流（BitTorrent）是一种内容分发协议，简要描述基于该协议的分布式内容共享系统的基本架构与工作方式。（6分）

混合式架构，集中式的文件目录服务，P2P的文件传输。

(3) 判定以下三项是属于“策略”还是“机制”。如果是策略，写出相应的机制；如果是机制，任意给出一个相应的策略。（6分）

a) 根据应用需求动态设定系统缓存的大小

策略

对应机制：可以改变系统缓存大小

b) 支持不同类型的加密系统

机制

对应策略：根据安全性要求，选择一种加密系统

c) 能够通过多条路径传送数据包

机制

对应策略：根据拥塞程度选择一条最合适的路径

(4) 简述进程与线程的异同；一个Web应用采用多线程结构有一定的优势，列举两个。（4分）

2. 通信机制（本题满分15分）

(1) 简述如何在TCP的基础上，实现远程过程调用（RPC）的“最多执行一次”（at most once）机制。（5分）

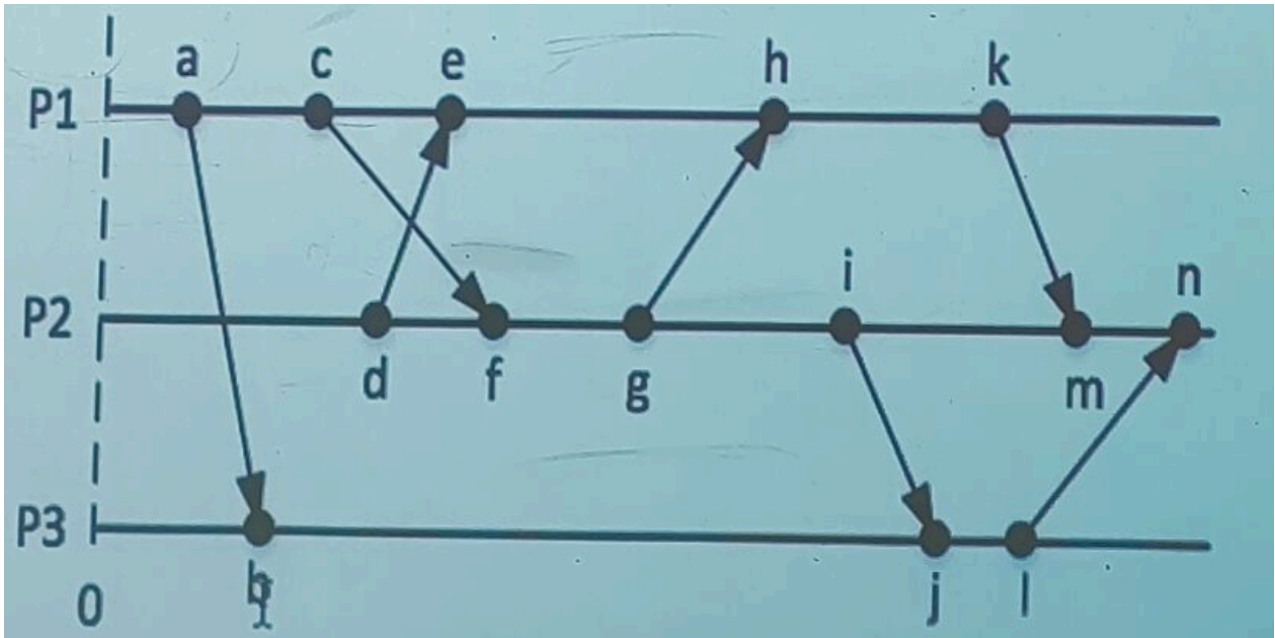
(2) 在不稳定的广域网上，不改变物理环境的前提下，请给出两个以上的方法来提升在线直播系统的流畅性，说明这些方法的基本工作原理。（5分）

(3) 电子邮件系统采用什么样的通信模式？简要说明如何在一般TCP/IP协议基础上实现这样的通信。（5分）

3. 同步（本题满分20分）

(1) 为什么在分布式系统中不适合用物理时钟进行同步？简述逻辑时钟的基本含义。（5分）

(2) 如下图所示，有P1、P2、P3三个进程以及发生的事件a、b、c、...，假定逻辑时钟的初始值为0。（15分）

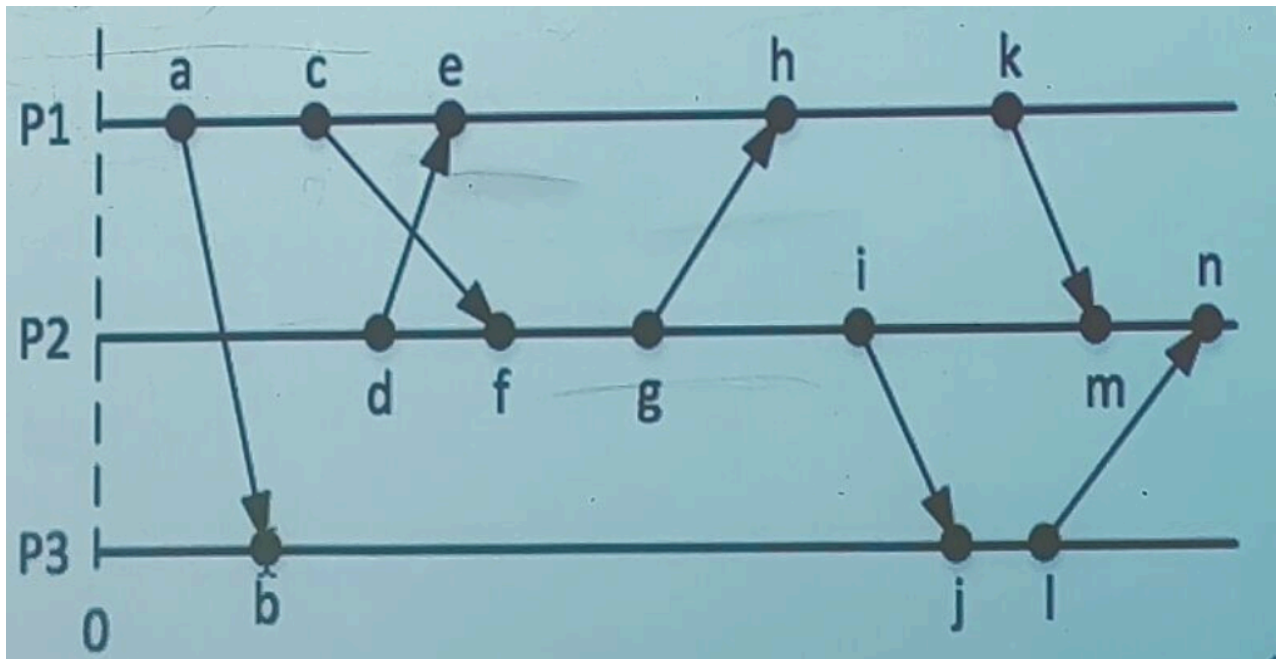


a) 为每个事件标定Lamport时间戳。（直接在图中标记）

b) 写出下面6组事件的先后顺序，如a发生在b前写作a->b，若无法判定则打X。

- (a, f) _____
- (d, h) _____
- (c, i) _____
- (g, k) _____
- (i, k) _____
- (j, n) _____

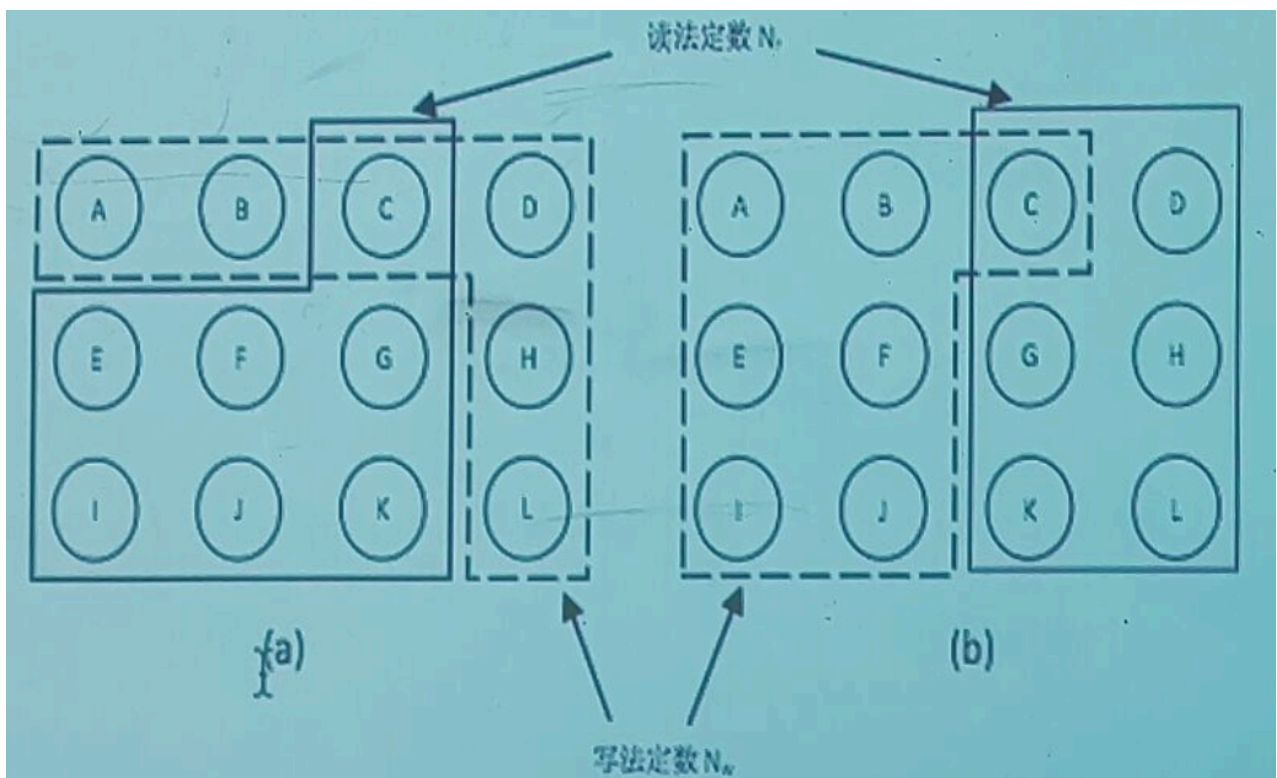
c) 设计任一时间戳格式，为每个事件标定全局逻辑时钟。简要说明方法，并在下图总标注。



4. 一致性（本题满分10分）

(1) 以“数据为中心”的一致性与“客户为中心”的一致性有什么异同？若希望通过增加副本来提升某门户网站的性能，副本可以采用哪种类型的一致性，给出具体的设计方案。（5分）

(2) 如图a和图b，是否符合法定数量（Quorum-based）协议的要求？为什么？（5分）



5. 容错性（本题满分15分）

(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。

需要 $(2K+1)$ 个进程。

(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，系统仍然能正常工作，总共需要安放多少个传感器？

需要 $(3K+1)$ 个进程。

6. 云计算与物联网（本题满分20分）

(1) 虚拟化技术在云计算中是如何运用的？如何理解部署基于虚拟化的私有云能提高现有系统性能？（6分）

(2) 大数据分析系统大多采用集中式的主从（master-slave）的架构，每个提交到系统的作业都会生成一个作业管理器来向系统的主节点索取资源。系统资源管理有两种方案：（6分）

一种是基于“拉”的方式（pull-based）：如Apache Hadoop YARN，每个运行着的作业管理器会定期（例如每3秒）向master报告心跳信息（索要资源），如果有资源的话，主节点向该作业管理器返回资源分配的结果。作业管理器拿到资源后就会在相应的节点运行其任务。

一种是基于“推”的方式（push-based）：如Apache Mesos，主节点发现有空闲资源时，会主动向作业管理器推送资源；若该作业想要这个资源，那么作业管理器就会发送作业到空闲资源上运行；若不要，则会给主节点发送拒绝消息，主节点再向其他作业管理器推送。

试分析这两种资源管理方案优缺点，并分别列举一个适合的应用场景。

(3) 在一个RFID系统中，阅读器通过识别其扫描范围内的标签获取感知数据。为了获取大规模的感知数据，现有 n 个阅读器（ n 远远小于标签总数），他们之间的扫描范围可能存在重叠，且少部分标签的位置会动态变化，如何选取最小数目的阅读器识别最大数目的标签？请提供一种集中式的解决方案和一种分布式的解决方案。（提供解决方案的主要步骤，无需具体计算）

注：1) 当多个阅读器同时扫描重叠区域的标签，将无法识别标签；2) 不存在某一个阅读器位于另一个阅读器扫描范围内的情况；3) 系统部署在二维空间中且阅读器的位置是已知的；4) 阅读器可以根据标签响应情况估算标签分布。（8分）

答案：选取最小数目的阅读器识别最大数目的标签，可转化为“带权最大独立集”问题。

集中式方案：

- 1) 将二维空间划分为多个子区域，如一个个格状区域；（1分）
- 2) 在每一个子区域中通过枚举求带权最大独立集；（2分）
- 3) 综合子区域的带权最大独立集，得到全局的带权最大独立集。（1分）

分布式方案：

- 1) 将二维空间划分为多个子区域，如一个个格状区域；（1分）
- 2) 在每一个子区域中，随机选择一个阅读器作为协调者，估算当前子区域的标签分布，并通过标签分布计算子区域中的近似带权最大独立集；（2分）
- 3) 综合子区域的带权最大独立集，得到全局的带权最大独立集。（1分）