#### State

#### 所有机器需要持久化的状态:

(在 RPC 响应之前,需要更新稳定存储介质)

**currentTerm** server 存储的最新任期(初始化为 0 且单调递增)

votedFor 当前任期接受到的选票的候选者 ID (初值为 null)

log[] 日志记录每条日记记录包含状态机的命令

和从 leader 接受到日志的任期。(索引初始化为 1)

# 所有机器的可变状态:

**commitIndex** 将被提交的日志记录的索引(初值为0且单调递增)

lastApplied 已经被提交到状态机的最后一个日志的索引(初值为

0 且单调递增)

#### leader 的可变状态:

(每次选举后重新初始化)

nextIndex[] 每台机器在数组占据一个元素,元素的值为下条发送

到该机器的日志索引 (初始值为 leader 最新一条日

志的索引 +1)

matchIndex[] 每台机器在数组中占据一个元素,元素的记录将要复

制给该机器日志的索引的。

## AppendEntries RPC

被 leader 用来复制日志 (5.3 节); 同时也被用作心跳 (5.2 节)

# Arguments:

term leader 任期

leaderId 用来 follower 重定向到 leader

prevLogIndex前继日志记录的索引prevLogItem前继日志的任期entries[]存储日志记录

leaderCommit leader 的 commitIndex

Results:

term 当前任期, leader 用来更新自己

success 如果 follower 包含索引为 prevLogIndex 和任期为

prevLogItem 的日志

### 接受者的实现:

1. 如果 leader 的任期小于自己的任期返回 false。(5.1)

2. 如果自己不存在索引、任期和 prevLogIndex、prevLogItem 匹配的日志返回 false。(5.3)

3. 如果存在一条日志索引和 prevLogladex 相等,<sup>是在新的entry中找是否有冲突的</sup>但是任期和 prevLogItem 不相同的日志, 需要删除这条日志及所有后继日志。(5.3)

4. 如果 leader 复制的日志本地没有,则直接追加存储。 <sup>改为强制截断</sup>

5. 如果 leaderCommit>commitIndex, 设置本地 commitIndex 为 leaderCommit 和最新日志索引中 较小的一个。

## RequestVote RPC

#### 被候选者用来收集选票:

#### Arguments:

term 候选者的任期 candidateId 候选者编号

lastLogIndex 候选者最后一条日志记录的索引

lastLogItem 候选者最后一条日志记录的索引的任期

Results:

term 当前任期,候选者用来更新自己 voteGranted 如果候选者当选则为 true。

#### 接受者的实现:

1. 如果 leader 的任期小于自己的任期返回 false。(5.1)

2. 如果本地 voteFor 为空,候选者日志和本地日志相同,

则投票给该候选者 (5.2 和 5.4) <sup>应该为"更新",在5.4节有定义</sup>

#### Rules for Servers

#### 所有机器:

1. 如果 commitIndex > lastApplied: 增加 lastApplied, 并将日志 log[lastApplied] 应用到状态机。 applyLogs

2. 如果 RPC 请求中或者响应中包含任期 T>currentTerm, 参与者 该机器转化为参与者。 每个RPC都需要首先包含这个规则

#### 参与者(5.2 节):

1. 响应来自候选者或者 leader 的请求。

appendEntries需要resetTicker

2. 如果选举定时器超时时,没有收到 leader 的的追加日志请求 或者没有投票给候选者,该机器转化为候选者。 RequestVote需要resetTicker

# 候选者 (5.2 节):

1. 一旦变为候选者,则开始启动选举:

1.1 增加 currentTerm

1.2 选举自己

1.3 重置选举定时器

1.4 并行发送选举请求到其他所有机器

, 新建gorou<mark>t i ne进行监听</mark>

合起来

2. 如果收到集群大多数机器的选票,则称为新的 leader。 根据测试文件,节点数必须大于1

3. 如果接受到新 leader 的追加日志请求,则转化为参与者。

4. 如果选举定时器超时,则重启选举。

leaders: 发送心跳时也可以顺便调整nextIndex

1. 一旦当选:发送空的追加日志请求(心跳)到其它所有机器;在空闲时间发送心跳,阻止其它机器的选举定时器超时。

2. 如果接受到来自客户端的请求,追加日志记录到本地日志。 如果成功应用日志记录到状态机则回应客户端。

3. 如果某个参与者的最新日志索引大于等于本地存储该参与者的 不要搞错成是 leader的 last log

最新日志索引:给该参与者发送包含从 nextIndex 开始的 Committegs 日志追加请求。

3.1 如果成功,更新该参与者的 nextIndex 和 matchIndex。(5.3 节)

3.2 如果由于日志不一致而失败,减少 nextIndex 并重试。(5.3 节) 4. 如果存在 N > commitIndex (本地待提交日志的索引),

majority(matchIndex[i]>= N)(如果参与者大多数的 最新日志的索引大于 N),并且这些参与者索引为 N 的日志的任期也等于 leader 的当前任期: commitIndex =N

(leader 的待提交的日志索引设置为 N) (5.2 和 5.4 节)。

updateCommitIndex