Action：

Agent负责调控每个调度周期内，切片组件的睡眠时长占比。动作空间中包含给定睡眠时长占比的调整步长[-50%, -25%, -5%, -1%, 0, 1%, 5%, 25%, 50%]。Agent选择合适步长调整到当前占比。

State：

为了让agent通过全局信息做出决策判断，我们用所有切片的流量负载和睡眠时间组成state：

Xxx

其中，V^n\_{i-1}，O^n\_{i-1}分别是上一调度周期内的切片$n$的总请求数据量和睡眠时间比例。

Reward：

切片$n$的reward设计为基站的总节能量，计算为：

Xxx，

其中，第一项是load-related的切片组件 saving power；第二项是基站整体deep sleep的saving power；第三项是切换切片mode的额外power开销，作为惩罚项。注意，x和x分别只包含切片n的对应能耗：xxx。

在每个调度周期的开始时间slot，所有切片的agent根据全局信息调整本次的休眠时间占比，以最大化系统整体的节能。多个切片组件的逐周期决策组成了基站睡眠控制的异构马尔可夫决策过程，接下来我们提出一个整体框架来求解此问题。