外 文 译 文

**第一章 绪论**

在家用无线路由器不熟密集的场景下，例如公寓楼，相邻的的家用无线路由器将在没有限制的情况下共享相同的频谱。在这种环境下，由于共享同一频谱的设备的密度的增加，无线路由器网路将会出现一些间断性的问题，比如无线丢包，来自无线网络及非无线网络的干扰。在这篇论文中，我们将提出一个厂商中立，基于云的集中式结构叫做COAP，通过一些植入这些无线路由器的一个公开的接口，来配置，整合，管理这些相互独立的无线路由器。这个系统，使用了OpenFlow的来进行了延伸，使得所有的无线路由器能够利用一个集中的控制器来共享各种的无线信息——无线干扰、流量情况、各个流的内容，来回复给路由器一些指导——配置参数信息（如信道）以及传输参数信息（通过粗糙的规划及参数的调节）。这篇论文将描述这个结构以及相关的技术，应用并展示其潜在的益处，例如，将我们实际部署在家用环境中的无线路由器的信道拥塞减少47%。

**第二章 集中式系统模型的动因**

当前密集的住宅无线路由器的部署使得我们在想要专门部署无线路由器的时候遇到很多问题。我们将简单的描述我们之前在部署30台无线家用路由器的时候所观察到的信息。

由于静态无线接入点的配置而导致的无线频谱低效的使用。在我们之前1个月的观察中，在300台无线路由器中有超过50%的家用无线路由器被静态的配置到一个单一的无线信道中。这就会导致这些无线路由器错过更多的机会来使用更好的无线信道来避免前面出现的问题。为了比较我们部署的30台在各个信道的相对活跃度，我们使用之前的记录来计算在部分时间内，在信道1、6、11中拥有最低的信道利用率的信道。这些数据，是从每个无线路由器的第二个无线网卡上获得的，这个网卡以一个轮训的方式，在1、6、11三个信道的跳转，中间的间隔是500毫秒。图2-1展示了，在一些地点，在各个信道的活跃度是十分不平均的（如无线路由器5，16，21）。在这些地点，有一个信道的最少会在80%的时间内被一直的使用，说明，其实这里面有更好的机会来进行更高效的信道设置。

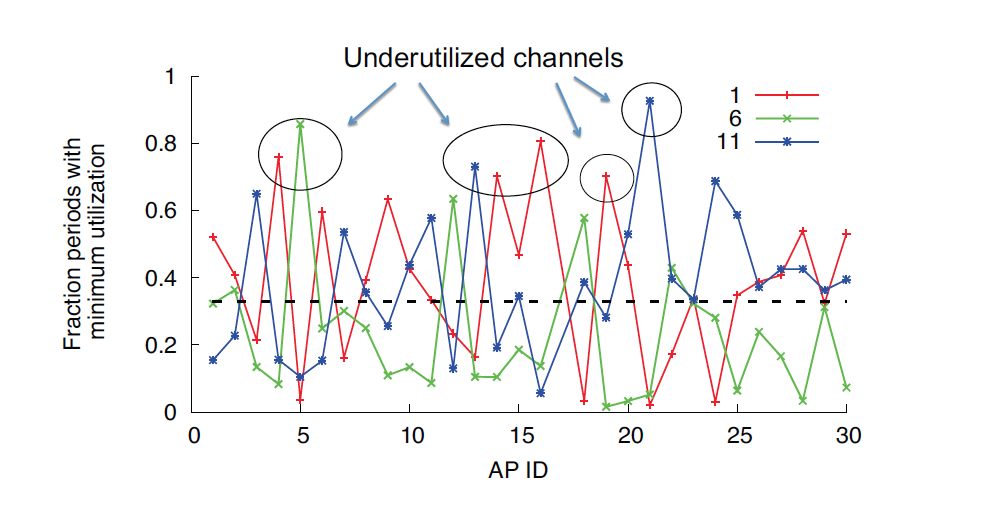


图 2-1 各个路由器信道利用率

无线路由器干扰。家用无线路由器由于一些无线参数，如过高的外部流量，客户端较低的物理速率，而会出现较高的信道利用率或严重的信道争夺情况。在一些用户流量使用活跃的时间段，周围无线路由器平均的信道占用率能够达到70%。因此某个无线路由器的一些活动就会进一步加剧周围使用同一无线信道的无线路由器体验到的信道拥塞情况。另外，有一些链接可能会偶尔经历周围的无线路由器带来的隐藏终端的干扰，而最终导致用户端的丢包。

非无线路由器干扰。不像802.11协议，有一些非无线路由器设备（例如微波炉，无线电话），当发现在同一信道有无线活动时，它并不会进行避让。由于这种冲突的无线发包，最终会导致接收端的丢包。在之前的文章中，作者观察到，在多个地点，微波炉能导致一些链接的性能80%的降低。

在以上的每一个例子中，如果一个无线路由器能够判定它的邻居无线路由器的无线内容（如流量信息）和无线信道，它就能够更好的决定它的补救策略，而不需要去完整的了解整个无线路由器的活动，因为每个无线路由器只能观测到其周围和它在同一信道的活动。如果周围所有的无线路由器能够在COAP的控制服务中心来整合这些无线信息，就能够进一步的指导这些涉及到的无线路由器进行可能的优化调整。

**第三章 通过COAP系统对于家用无线路由器进行杠杆平衡调控**

为了从这个系统中的到最大的收益，在一个多房间的住宅楼里，可以使用一个简单的控制器来管理这些COAP无线路由器。这里我们将讨论其在信道分配上面的使用。

**3.1 无线信道的配置**

在企业无线路由器中，它的配置（如信道，传输信号强度）是由设备厂商一个特定的集中控制器来管理的。而在家庭使用中，这些配置往往是有个人手动执行或基于当前无线路由器的观测状况完成的。这就会导致前面提到的对于无线频谱的不充分的使用。

在COAP这个系统中，这些的功能（如信道分配）就可以被交给控制器来完成。例如，通过整合从COAP的无线路由器的周围获取到的最近信道利用率的信息（如在最近的x分钟内），相比于只根据自己本机的观察结果，这个控制器就能够更好的完成对于信道的分配。为了更好的实现这个应用，我们首次分析了整合周围无线路由器信道利用率的信息，来增大本机无线路由器对于无线频谱的观测。

这个控制器能否在空间及时间上更好的调整周围的无线路由器，来达到一个更好的信道选择吗？

对于家用无线路由器，它只能观测到它当前所在的无线信道的活动信息。这可能就会造成这个无线路由器在信道拥塞的情况下，失去选择一个更好的信道的机会。利用COAP，通过周围无线路由器所观察到的无线活动情况，那一个无线路由器所能观测到的整个无线频谱的信息就会增大。我们研究了通过周围无线路由器（接收信号强度大于-55dbm）信息在其他信道上的信息，来估计其他信道的信道利用率的可行性，这我们将在下面讨论。

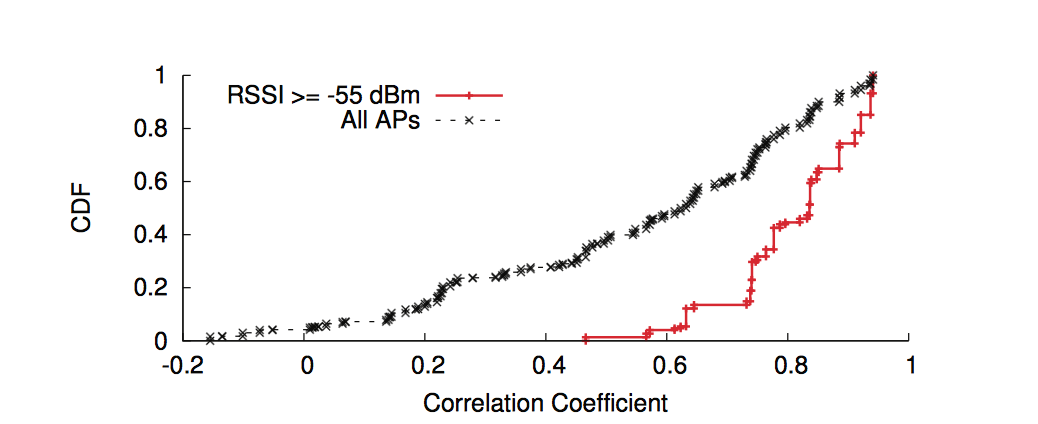


图 3-1 信道利用率之间皮尔逊相关系数

利用我们在之前中获得的记录信息，图3-1展示了周围的无线路由器对在每个信道上在不同的时间序列间（每分钟平均）的信道利用率之间的皮尔逊相关系数的累积分布函数。从这幅图中可以看出，有超过60%的临近无线路由器对，当他们的接收信号强度大于-55dbm时，就会表现出皮尔逊相关系数大于0.8的高相关性。另外，几乎所有的无线路由器都会有高于0.6的相关性系数。对比之下，在所有的无线路由器对中，在这些的信道利用率中，其中只有15%会表现出超过0.8的相关性。这些结果说明，通过周围路由器探测周围在时间及空间上的无线活动，来估计其他信道上的信道利用率，是可行的做法。这个可以帮助COAP的控制器，通过利用一些著名的算法来实现更好的信道分配。

集中式的信道分配能够带来哪些性能上的提升？

为了更好的研究COAP系统对于信道分配的好处，我们利用我们部署的12台相邻的COAP无线路由器来进行了，基于信道利用率的动态信道分配，来尽可能的减少信道利用率及信道拥塞来最终提升吞吐率。由于我们无线路由器的部署稀疏，我们使用在我们的无线路由器上的第二块无线网卡来对所有的信道，进行一个轮训的信道利用率信息的收集。在实现时，我们队周围路由器上的流量进行了模拟，为了我们收集信道利用率的信息。

在6天的采集中，我们交替的部署了2种的无线信道的配置策略：静态（或随机）的信道分配，通过COAP动态的进行信道分配来减少COAP无线路由器的信道利用率。在第二种策略中，我们叫使用从最近5分钟内采集的信道利用率信息来估计外部无线活动，来最终为每个无线路由器选择一个最好的信道。为了避免对用户的使用产生影响，这些信道的配置更新只有在平静时刻才会被应用。

图3-2展示了使用了基于COAP的动态信道分配的12台COAP无线路由器，其对比只用静态信道分配策略的路由器的信道利用率的中位数及90分位数的减少的百分比。这个说明，对于12个无线路由器中的10个来说，动态的基于信道利用率的信道分配要好于随机的信道分配。对于这12个中的4台无线路由器，都有20%甚至超过47%的信道利用率的减少。这也就体现了通过集中式的方法来进行家用无线路由器管理所带来的潜在的好处。这样，即便是没有参与到COAP的周边的无线路由器也能收获到其带来的好处，因为整体信道拥塞的降低。并且目前日益增多的无线设备及高流量，也将会进一步扩大基于COAP的方法来减少信道拥塞的益处。

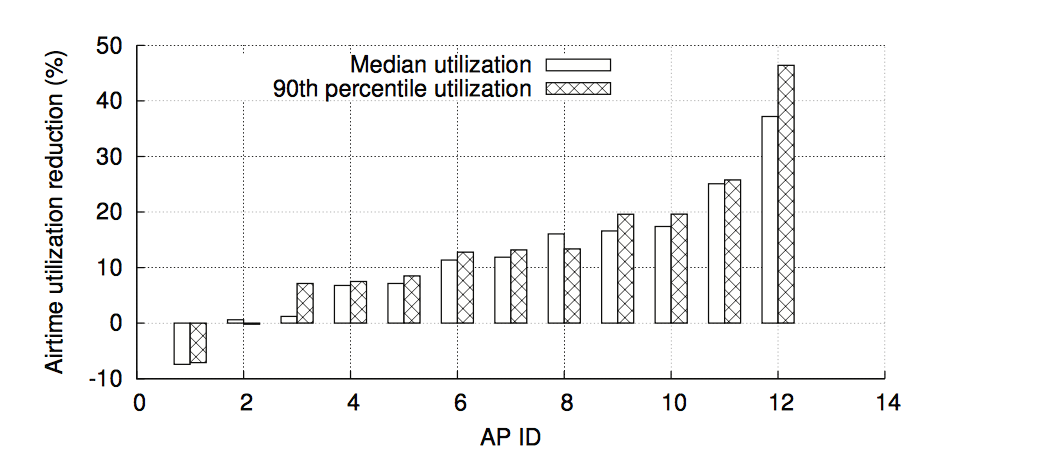


图 3-2 动态信道分配后信道利用率减少比率

**第四章 学习之前的内容信息来预测未来的活动**

我们的控制器能够分析之前从COAP的无线路由器中获取的无线活动信息，来使用学习技术和挖掘“内容相关”的信息。这里我们将在这章中提出一个例子：利用用户设备类别及流量信息来预测流量特征。

**4.1 通过学习用户设备及流量信息来预测未来的活动**

由于家用无线路由器对于所有的无线设备来说起到的是一个网关的作用，因此他们可以捕捉到所有有用的粗糙的有关无线活动的内容信息。这包括用户设备类型信息（如iphone，google电视，平板）以及流量来源的id信息（如Netflix, Pandora）。在COAP的架构中，控制器可以利用这些之前的信息来预测未来流量的特征，以及对于无线路由器采取基于流量信息的配置。下面是一个场景实例——无线路由器4，在它的无线电视上开始了一段高流量的高清视屏，它有90%的可能性会持续超过30分钟，因此可以将周围的无线路由器移到其他的无线信道上来减少在这个流量过程中的信道冲突。

设备类型和流量内容有什么帮助？

在下面的分析中，我们将使用从部署的COAP的无线路由器获取流量的信息，来挖掘设备类型和流量内容对于流量特性预测的帮助。流量的来源将由与每条流量对应的顶级域名确定。对于我们部署的COAP无线路由器，这个顶级域名是可以被哈希的并且是匿名的，以此来保护用户的隐私。用户设备类型是由其mac地址的前24个比特（这个将决定组织的唯一标识符）以及平衡了从主机名（例如很多的设备，比如Iphone，就会将设备类型写入主机名中）来决定的。我们使用这些流量信息来抽离出这些流中的“活跃时间段”（每秒大于40个包），并且获取了在这些活跃时间段中的流量特征信息。

突发性属性：一次突发是指在一段时间内，在持续的活跃时间段间的间距小于10秒。对于这些突发性的流量，我们将计算经历的时间，下载的比特数，平均以及最大的下载速率。

会话属性：一次会话是指在一系列连续的流量突发之间的间距小于5分钟。这个过程表明，其正在活跃的使用流量源头的信息。对于每次会话，我们将计算在持续的突发间的间隔时间，整个会话的持续时间，下载的比特数，以及下载速率。

通过流量类型获取对于流量使用的偏向性。在我们的记录中，Netflix是产生流量最大的流量来源。图4-1中显示了，排名靠前的Netflix用户设备占Netflix流量的百分比。通过对于每种设备的分布分析，我们可以看到用户设备和流量使用的一个紧密的相关性。而这种倾向性就可以被我们用来预测流量的特征。

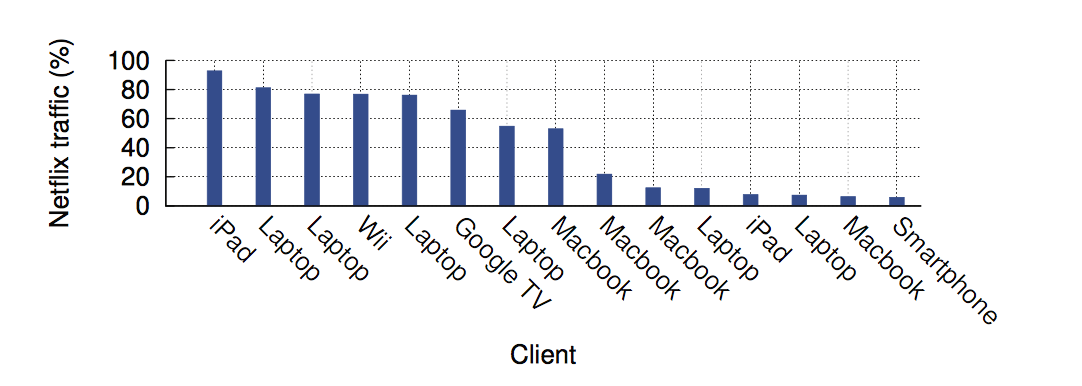


图 4-1 Netflix用户占比

设备使用流量特性的影响。设备类型也同样可以影响用户行为（比如他们花在这个设备上的时间）。例如，一个用户可能只会使用他的智能手机很少的一段时间，但是却会使用他的游戏机很长的一段时间。由图4-2可以看出，任天堂设备有着更高的会话时间（中值是20分钟），而其他的设备，比如笔记本电脑及iPhones，他们中值只有5分钟和2分钟。

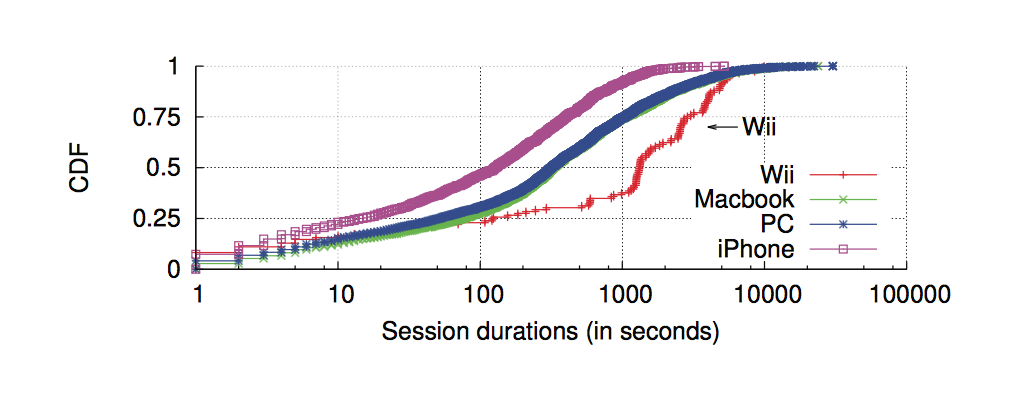


图 4-2 各种用户设备会话时间分布

不同平台下的流量特性。使用平台的类型（如用户设备，操作系统）也会影响流量特性。例如，视频服务提供商，就会针对不同的浏览器，设备类型，采取不同的策略。在我们的分析中，我们发现对于一些十分流行的视频来源（如Netflix）有着一些相似的特征。图4-3告诉我们，对于任天堂的设备，它所能在每个突发流量达到的最大的下载速率只有5Mbps，但是对于笔记本电脑以及Google电视，最大值却能达到16Mbps。由于我们并没有这些视频流量的更多的信息，但是我们相信，对于任天堂设备的较低的速率，是可能由于Netflix应用对于这些设备的一种速率限制。

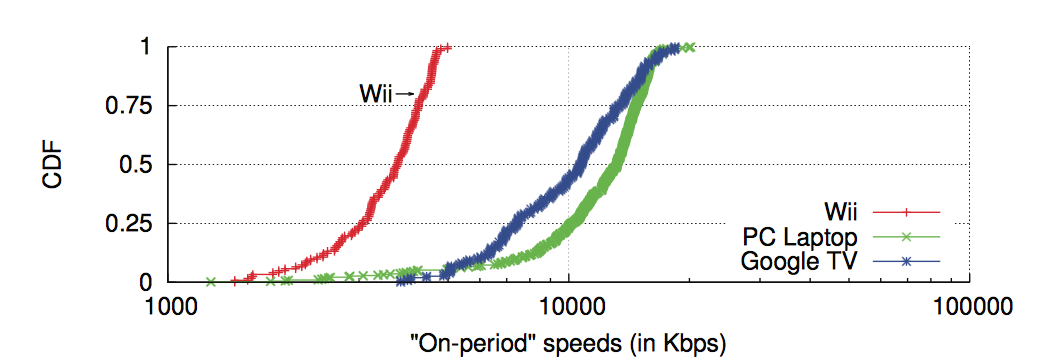


图 4-3 三种设备达到的速率分布

**4.2 预测流量特性**

为了将通过用户设备及流量信息来预测前面提到的流量属性（如会话时间，下载比特数）所带来的帮助进行一定的量化，我们采用了机器学习的工具，Weka，以及从COAP的无线路由器采集到的流量信息。

我们在这里用“相关信息”来代表，我们采集到的以下的和流量和设备相关的特性信息——无线路由器的ID，用户设备ID，流量来源ID，时刻，星期。这些都是从一个新的流量一开始就可以获得的信息。这些输入的特征就可以用来预测与突发性流量及会话流量相关的一些特性（如会话长度，在每个会话中下载的比特数）。我们对比了当只使用无线路由器ID，时刻及星期作为输入参数（不包含任何流量和用户设备信息）及使用所有参数的预测效果。对于输入的数据及评价的数据我们采用了10折交叉评价法。

预测效果。我们尝试了Weka中的多种不同的算法，最终发现REPTree的算法的效果是最好的。它使用了信息增益作为建立这个树的模型的基础。

图4-4的上半部分及下半部分，分别展示了对于每个无线路由器及每个用户设备类型（如任天堂，Iphone，笔记本电脑），针对会话长度以及每次会话中下载的比特数的预测错误率的累积分布函数。它表明了，使用用户设备和流量信息，可以将预测的错误率降低2倍，甚至更多。并且对于用用户设备及流量信息的80分位的相对预测错误率都低于100%说明，大部分情况下，是有可能预测这些流量属性的。同时我们也观察到了在预测其他流量属性（如下载速率）时，也会有的类似的结果。因此，用户设备类型及流量信息可以向COAP的控制器提供有价值的信息来帮助它对于未来可能的无线活动进行预测，并根据这个来执行可能的优化措施。

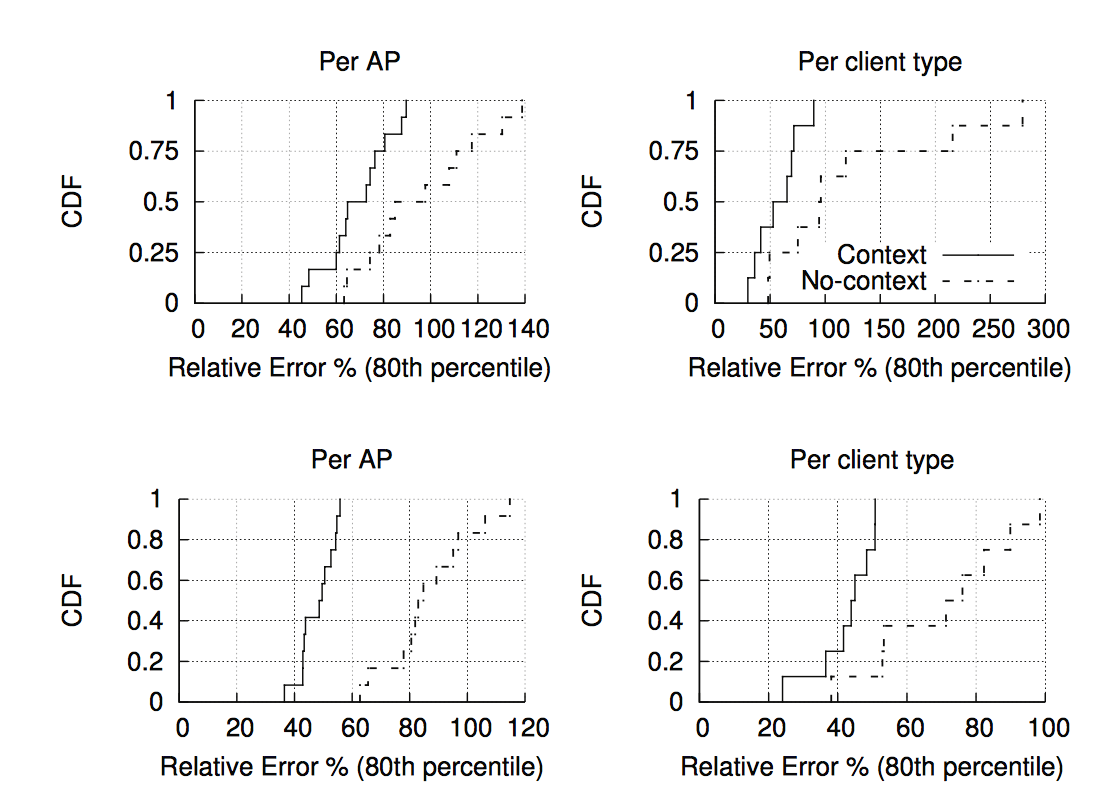


图 4-3 机器学习预测错误率分布

**第五章 总结**

我们提出了COAP系统，它采用了设备厂商中立的一个开放的API，以及一个基于云的中心控制器，最终能够使得多种密集部署的家用无线路由器能够协同工作。我们描述了这种集中式的系统的多种应用场景——利用从周围无线路由器获得的信道利用率信息来进行更好的信道分配，通过管理周围无线路由器对于信道的使用，来减少对于重要的流量（如视频流量中的HTTP）的拥塞干扰，以及减少隐藏终端的干扰。已经学习之前的无线活动信息来使得中心控制器能够更好的预测未来的无线流量特性。这个也能够使得控制器能够提前的对于家用无线路由器进行配置，来减轻周围无线路由器所带来的干扰，并且能够基于流量来给出相应的优化建议。