

无线局域网中的负载均衡技术

Load Balancing Technologies in WLAN

王彬/WANG Bin

(中兴通讯股份有限公司 中心研究院西安研究所, 陕西 西安 710065)
(Institute of Xi'an, Central Academy of ZTE Corporation, Xi'an 710065, China)

中图分类号: TN929.5 文献标识码: A
文章编号: 1009-6868 (2006) 03-0046-05

摘要: 随着无线局域网扩展服务集中的终端(STA)的移动和无线信道的时变, 各个接入点(AP)上的负载会出现差异, 需要加强负载均衡, 平衡各个AP的差异, 以达到最大化的网络资源利用。无线局域网中的负载均衡算法可以采用STA/AP主控负载均衡或接入式/切换式负载均衡技术。在时分双工的时分同步码分多址(TD-SCDMA)系统中, 由于系统在最小资源单位和媒体访问机制等方面的差异, 负载均衡可以在小到时隙和码道, 大到终端和小区等不同的级别上进行, 体现码分多址(CDMA)和时分双工(TDD)的特点, 具体方式更加灵活多样。

关键词: 无线局域网; 负载均衡; 算法

Abstract: Because of the movement of Stations (STAs) and the time-variant characteristic of wireless channels in the Extended Service Set (ESS) of Wireless Local Area Network (WLAN), each AP's load is different. Load balancing needs to be strengthened to balance the diversity of Access Points (APs), so as to use network resources efficiently. The load-balancing algorithm in WLAN may use technologies such as STA/AP host-control load balancing or access/switch load balancing. In a TD-SCDMA system, its minimum resources unit and media access mechanism are different from others, so the load balancing can be realized at different levels: based on time-slot and code-channel, or based on user ends and cells. With the characteristics of CDMA and Time Division Duplex (TDD), the load balancing can be more flexibly implemented.

Key words: wireless local area network; load balancing; algorithm

在有线分布式系统中, 负载均衡是指将网络的外部服务请求均匀分配到多台后端的服务器上, 从而提高网络服务质量的方法。由于任务到达的随机性, 以及各处理节点处理能力上的差异, 当分布式系统运行一段时间后, 某些节点分配的任务还很多(称之为超载), 而另一些节点却是空闲的(称之为轻载)。一方面, 使超载节点上的任务尽可能快地完成是当务之急; 另一方面, 某些节点轻载是一种浪费。如何避免这种忙与空闲并存情况的发生, 从而有效地提高系统的资源利用率, 减少任务的平均响应时间。这是负载均衡需要解决的问题。

有线分布式系统中的负载均衡方法操作的对象是任务。收集负载信息, 进行相关算法决策, 对分配给各节点的任务进行重新调度如进程迁移或任务迁移, 是负载均衡的基本途径。负载均衡问题是对任务的优化组合问题。

负载均衡技术应用到无线系统下具有新的特点。无线局域网中的负载均衡面临比有线网络更多的困难。随着无线局域网中扩展服务集中的终端的移动和无线信道的时变, 各个接入点(AP)上的负载出现差异。负载均衡的作用就是平衡各个AP的差异, 以达到最大化的网络资源利用。负载均衡是无线资源管理的重点, IEEE的802.11工作组的TGv任务组, 专门负责解决无线局域网的负载均衡问题。

1 有线分布式系统中的负载均衡

1.1 有线分布式系统负载均衡的技术难点

负载均衡需要解决的核心问题有两个: 一是何时进行任务迁移, 另一个是怎样进行任务迁移。要实现负载均衡, 难点在以下几个方面:

- 对系统的负载情况进行准确的评价。
- 负载均衡要求进行某些额外附加处理, 如系统内各节点负载信息的收集、存储、决策、任务迁移等。这些额外处理带来额外的网络和主机开销, 是对系统性能的损害。
- 要避免负载抖动, 即被迁移的任务在各节点之间被不断迁移而得不到执行。
- 解决系统各节点的硬件或软件的配置差异问题。

当前有线网络的负载均衡集中在Web服务器、邮件服务器、电子商务中的业务服务器等环境下的应用问题。

1.2 有线分布式系统中负载均衡算法的分类

负载均衡有许多种分类方法, 但是从整体上可以按照与系统的负载状态是否相关, 分为状态无关均衡和状态相关均衡两大类。前者又叫静态均衡, 后者又叫动态均衡。静态负载均衡是机械式的, 按固定的模式或规则来分配客户请求, 而不管服务器的运行情况。动态负载均衡则带有智能化的特征, 能够根据各台服务器运行时负载情况来恰当地分配请求。

状态无关均衡, 就是根据以往的经验或系统本身信息的收集, 把外来的任务分配给各个节点, 或对某些节点上的任务进行重新分配。由于这样的均衡决策是与系统当前状态无关的, 带有一定的盲目性, 因而, 决策的准确性很低, 通用性差, 有时甚至会造成与预期完全相反的结果, 加剧不均衡的程度, 造成系统性能的恶化。

状态相关均衡, 其决策取决于系统当前的状态, 也就是说, 系统可以根据当前的负载分布情况, 对各个节点上的任务进行动态的调整, 使已经分配给超载节点上的任务, 通过通信设备, 迁移到轻载的节点上去, 从而提高系统的资源利用率, 减小任务的平均响应时间。动态均衡法通常包含以下3个方面:

(1)各个节点间负载信息的收集, 方法通常有3种: 周期性收集策略、命令驱动策略、状态变化驱动策略。

(2)根据所收集的负载信息进行决策: 节点根据收集的负载信息, 确定需要进行负载迁移的节点对。发生负载迁移的节点, 根据实际需要, 可以有一对或多对。

(3)实现任务在各节点之间的迁移: 在负载均衡系统中的任务迁移可分抢先任务迁移和非抢先任务迁移两种。

2 无线局域网中的负载均衡

无线局域网中的负载均衡与有线分布式系统中的负载均衡不同, 表现在负载均衡控制的对象和控制方式上。由于无线终端的移动性和无线信道的时变性, 无线局域网中的负载均衡面临的情况更加复杂。

2.1 无线局域网的特点

无线局域网(WLAN)是指应用无线通信技术将计算机设备互联起来, 构成的可以互相通信和实现资源共享的网络体系。《ANSI/IEEE Std 802.11, 1999 Edition》^[1] 是IEEE 802.11无线局域网标准工作组制订的物理层和媒体访问控制协议标准, 便于无线局域网运营商及无线设备制造商建立互操作网络设备。物理层定义了数据传输的信号特征和调制方式, 媒体访问控制层可视为由一系列可完成诸如信息交换、功率控制、关联管理、同步管理和过程管理的服务组成。

在无线局域网中有两种角色: AP和终端(STA)。AP作为固定基础设施, 通过有线分布式系统互连。STA作为移动终端, 接受AP提供的接入服务。

WLAN共有两种工作组织形式, 独立基本服务集(IBSS)和基本服务集(BSS)^[2]。IBSS表示的就是Ad hoc方式的终端独立组织成一个网络单元, 在这种方式下没有固定的基础设施, 所有的终端直接与其他终端互连, 网络组织非常独立。BSS表示的就是基础模式, 至少一个AP作为网络的管理设备, 管理所有的STA, 向STA提供接入和传输服务。为了扩展无线局域网的覆盖范围, 增加终端的移动性, 多个BSS可

以联合组成扩展服务集(ESS)。扩展服务集通过扩展服务集识别号(ESSID)来区分, AP通过基本服务集识别号(BSSID)来区分。多个AP工作在不同的无线信道, 通过有线分布式系统互连。终端可以在ESS所覆盖的范围内自由移动, 并随着移动从ESS内的一个AP切换到另一个AP。

无线局域网的网络结构如图1所示。

终端通过扫描信道, 获取所处WLAN环境中扩展服务集和基本服务集信息, 并选择加入一个合适的基本服务集。WLAN中负载均衡就是由多个AP为多个移动终端提供接入服务的系统, 每一个AP通过利用网络信息进行站点管理, 目的是充分利用网络现有资源, 减少网络阻塞。

图1中, 位于中心点的AP, 经过一段时间的运行, 连接的终端的个数要多于四周AP的终端个数, 承担的业务也已经饱和, 此时连接到中心点AP的终端的业务的服务质量(QoS)已经不能保证。而外围的AP业务还不饱和, 网络中出现负载差异。负载均衡就是要解决这种差异, 将中心点AP的某些业务转移到外围的AP上, 以提高网络的服务容量和中心点AP的业务质量。

STA在进入网络的时候必须选择一个AP作为连接对象, 这就是终端接入过程, 接入过程要经过扫描、认证、连接等步骤。一般而言, 在ESS中, 终端在扫描过程中会扫描到多个AP的信息, 终端选择AP的条件是扫描到的AP的信号强度。由于无线局域网中STA的移动特性和无线信道的时变特性, STA在某些情况下会发生从一个BSS切换到另一个BSS的情况。这种情况就是无线局域网中的切换。终端切换带来的问题是切换时延, 甚至上层业务的中断。切换还会带来的问题是“乒乓效应”。即STA不停地在两个或多个AP之间切换, STA就像乒乓球一样。这样会造成负载抖动, 对网络资源造成极大的损耗。

2.2 无线局域网负载均衡的技术难点

当前的IEEE 802.11标准并没有对负载均衡方面做出描

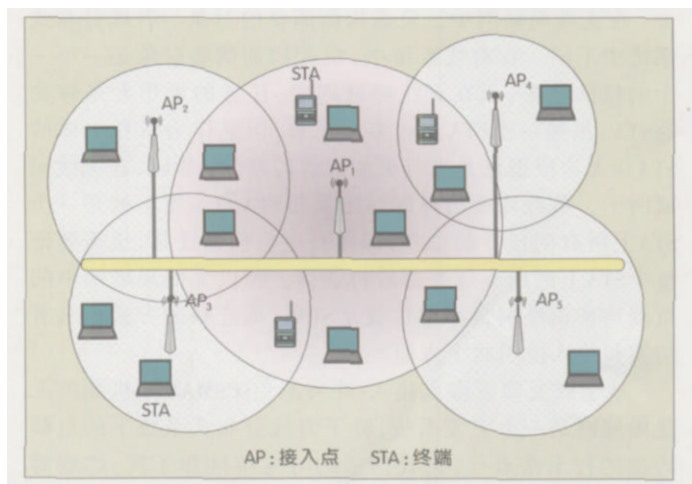


图1 无线局域网的网络结构

述和规定,网络负载的分布完全靠终端的业务自由和随意地切换决定。没有负载均衡的系统的缺点表现在以下几个方面:

(1)无线网络的效率和性能会降低

扩展服务集作为一个整体向终端提供服务。通常无线网络中的热点出现后,某AP的负载可能超过限制,大量的业务和用户竞争该AP的资源,造成该AP上的业务的QoS发生恶化,并可能长期处于这种状态。如果其邻接AP的负载水平却相对较低,没有负载均衡调整的话,整个网络将处于一种非稳定状态,无线网络的效率和性能都会因此而受到影响。

(2)增加网络拥塞的风险

没有负载均衡控制的网络,终端的切换是自由和盲目的,随着热点的出现,大量的终端会涌向热点AP,离开原有的AP。而终端向目标AP切换的结果是不能保证的,当热点AP的业务饱和之后,新的终端就不能接入,或者虽然接入,但是造成AP的拥塞,所有的业务中断。

(3)降低网络的容忍度

同样,当某一AP的负载饱和后,新的终端的接入会造成严重的后果,但此时,网络中其他非饱和AP还可以承接终端的接入,向更多的终端和业务提供服务。如果采取负载均衡调整,负载饱和AP的任务就可以被分担,整个扩展服务集中的AP的负载保持相对稳定和平衡,可以容忍网络的波动和变化。

(4)资源利用率差

没有负载均衡的网络,不能统筹调度和安排网络负载的分布,往往随着某一个或几个AP的饱和出现饱和,而其他非饱和的AP的资源无法得以利用。

总之,没有负载均衡的网络,业务的发生处于自由而不受控的状态,网络的稳定性很差,随着热点的出现,很容易出现终端拥塞,导致局部区域的网络瘫痪,或者终端业务的QoS指标降低。

在无线局域网中,负载均衡调整的对象与有线分布式系统中不同。在有线系统中,负载均衡调整对象是一个一个的排队任务,而在无线局域网中,任务的承担者是移动的STA,虽然一个STA上有多个任务,但是作为单独个体的STA却无法根据任务进行更加细致的划分。所以,在无线局域网中,负载均衡的控制对象就是STA,实质上就是一个STA上所有的任务的总和,即便是进行任务迁移,也需要把整个STA上的所有任务都进行迁移。所以无线局域网中的负载均衡调整的最小单位就是STA,而迁移的方式也只有切换和接入控制两个方面。

基于载波侦听多点接入/冲突避免(CSMA/CA)机制的无线局域网是一个突发系统,除了有线分布式系统下的负载均衡的技术难点外,无线信道的时变特性和干扰,终端移动性等会引入更多新的技术难度:

(1)负载信息的收集困难

在扩展服务集中,多个AP工作在不同的信道上,虽然可以通过AP之间的分配系统(DS)进行传输,但属于设备间的私有协议。IEEE 802.11k致力于解决无线网络测量问题,对于负载信息而言,最优的测量者还是AP。但是负载信息的交互却没有好的方法。另外一个关于负载信息的困难是对负载信息的量化困难。对于分组交换的无线网络而言,除了业务量之外,业务的QoS要求也是很重要的一个方面。终端数量将影响网络管理的开销,也是负载信息中的重要部分,还有一些其他因素如发射功率、信道选择等,都需要在评价AP的负载水平时加以考虑。

(2)决策过程的困难

由于终端是可以移动的,而终端上的任务是随机突发的,所以在一个时刻收集到的负载信息,对未来的均衡调整策略的真实有效性是需要评估的。在决策的过程中要充分考虑到这些因素。另外,负载均衡的调整目标的选择是很困难的。在无线网络决策确定了负载均衡的目标之后,参与者要通过切换将负载转移。终端进行一次切换的时间是100 ms以上,即便是快速切换也需要60 ms左右的时延。而且一次调整的对象也不仅仅是一个,所以,在这段时间之内,网络状况有可能已经发生变化。负载均衡的决策算法还涉及到降低参与者以减少管理开销,避免造成网络中大的波动等问题。

(3)任务迁移的困难

任务迁移是通过终端的切换实现的,而终端切换首先是引入时延,对终端的相关业务造成影响。其次会引入网络管理的开销,增加的开销本身就会对AP造成管理负担,对整个网络的性能也是一种影响。另外,在无线局域网中的负载均衡必须考虑STA切换失败的风险,因为虽然AP可以做出任务迁移或切换的决策,但是终端不一定能够连接到负载轻的BSS上,造成的结果就是终端离开后,不得不重新寻找新的接入点,上层业务同时中断。负载均衡的收敛与否也和终端切换的结果相关。

负载均衡已经被作为IEEE 802.11v中的主要工作组内容,纳入了标准的制订工作。快速切换是IEEE 802.11r标准的工作内容。快速切换和负载均衡内容是紧密联系的。

2.3 无线局域网中的负载均衡的分类

可以从两个方面对无线局域网的负载均衡技术进行分类。按照负载均衡的发起者分,负载均衡可以分为STA主控的负载均衡和AP主控的负载均衡。按照负载均衡的解决方式可分为接入式负载均衡和切换式负载均衡。无线局域网的负载均衡的分类如图2所示。

2.3.1 STA主控负载均衡

STA主控的负载均衡实际上是一种自发的负载均衡策略,是无线局域网中STA的特性之一。切换完成的STA实际

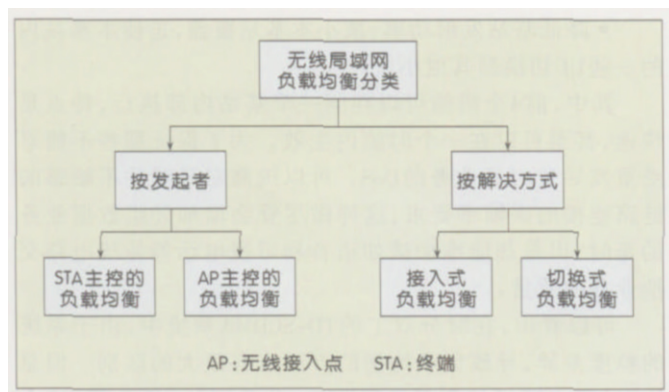


图2 无线局域网的负载均衡分类

上也实现了网络负载的传递。

仅仅靠信号质量作为STA切换的条件是片面的，目前有大量的研究集中在如何改进切换的时机，以更准确地实现终端切换带来的负载变化。如UT斯达康通信有限公司在2004年4月27日申请的专利《无线接入点负载均衡方法》(专利号: CN200410017978)中，对终端切换的条件加以扩充，在终端接入无线接入点之前，比较ESS内AP的当前接入用户数或接入数据流量，无线终端接入到接入用户数或接入流量最少的无线接入点，从而实现局域网无线接入点的负载均衡。终端接入或切换除了AP的信号强度，还要考虑AP的当前用户数和流量，增加了切换目标AP选择的准确性。这样的方法的缺点是切换或接入的过程中，终端必须查询AP的状态，从而增加了切换的工作步骤，使切换的时间延长，违背快速切换原则。有研究改进了此方法，由AP定期广播其流量和用户数量，这样做虽然可以在终端背景扫描时就一并得到信号质量和负载水平，但是会增加网络开销。

STA主控的优点是: STA最清楚自己的业务需求和所处的环境，STA扫描得到的环境中可用的BSS情况对STA而言最准确。而且，切换或接入的主体是STA，业务也是在STA上开展，切换造成的损失由STA承担。缺点是: STA仅仅知道终端侧的情况和STA自身的情况，对网络环境而言，STA并不能准确掌握，没有发言权。STA切换随意性大，会对整个ESS造成混乱，使网络全局的负载均衡效果不佳。

2.3.2 AP主控负载均衡

AP主控负载均衡的出发点是，AP处于网络侧，AP之间通过有线的连接，可以在不影响STA业务的情况下，相互传递与负载均衡相关的信息。根据这些信息，ESS就可以判断出AP中负载最重的AP和负载最轻的AP。拥有这些信息的ESS就可以做出一些决策，比如，让负载重的AP下的终端切换到负载轻的AP下，或者对新加入网络的终端进行限制，不允许负载重的AP接入而只允许负载轻的AP接入等，实现ESS内的负载均衡。

目前的研究重点集中在AP主控的负载均衡方法上。如

联想(北京)有限公司在2003年9月24日申请的专利《一种无线接入点的负载均衡方法》(专利号: CN200303134660)中采用以下方法实现负载均衡: 通过无线接入点之间的相互通信，每个无线接入点可以根据当前网络的负载分配情况决定是否允许新终端接入，如果负载高，则拒绝终端的接入，从而实现负载均衡。IBM在2002年6月26日申请的专利《Access Point Initiated Forced Roaming Based Upon Bandwidth》(专利号: US20040001467)中提到一种控制终端移动和漫游的负载均衡方法: 无线接入点监控其管辖的所有终端的带宽使用情况，如果发现负载异常，就选择一个终端，强迫其漫游到其他的无线接入点上，由此实现负载均衡。

中兴通讯股份有限公司也在这方面有深入的研究，并有大量的成果。如王占利在2004年3月10日申请的《一种无线局域网接入点设备间负载均衡的实现方法》(专利号: CN200410008410)中采用以下方法实现负载均衡: 多个接入点之间通过交互消息的方式选举出主控接入点，主控接入点允许移动无线终端接入，其余接入点设为备选接入点，备选接入点拒绝移动无线终端接入。这个方法可以克服负载均衡的实现依赖于接入点AP外部服务控制器的控制或只能针对特定STA用户群应用的缺陷，解决了接入点AP间充分利用有效资源实现动态负载均衡的问题。

AP主控的优点是网络侧的整体情况可以通过AP之间的分布式系统，准确、及时地进行统计，在整体协调和调度方面对网络侧有利，也比较容易实现整体策略的实施。另外就是整个网络的负载均衡调整速度会很快。缺点是网络侧并不知道STA所处的位置和网络环境、STA所能够扫描到的AP的情况，所以有切换决策错误的风险。

2.3.3 接入式负载均衡

接入式负载均衡就是控制STA的接入实现负载均衡。当AP的负载情况超过阈值后，该AP就会拒绝新的终端的接入，加入网络的终端只好寻找负载较轻的AP进行连接。从而在一定的程度上实现负载的均衡。

由于接入式负载均衡只控制新终端的接入过程或切换后的重连接过程，属于被动的负载均衡，负载均衡的调整收敛速度会比较慢。极端情况下，没有新的终端进入网络，或者所有的STA都维持当前连接，网络会一直处于非平衡状态。

2.3.4 切换式负载均衡

切换式负载均衡就是控制STA的切换以实现负载均衡。当ESS中某AP的负载过重的时候，ESS或终端控制该AP下的STA切换到其他AP上，以分散负担。当ESS中某AP的负载过轻的时候，ESS或终端就控制其他AP下的STA切换到该AP上，以实现整个ESS的全局平衡。

切换式负载均衡的特点是反应迅速，一旦出现负载差异，可以通过切换行为迅速调整。缺点是切换会对终端造成损失，有切换失败和上层业务中断的风险。同时，切换式

负载均衡必须依赖于快速切换才能实现。如果快速切换方法效果差,则负载均衡造成的损失会增大。

将AP主控方法和切换式负载均衡方法相结合是当前研究的热点,不仅可以准确地收集信息做出决策,而且可以迅速反应,主动地调整扩展服务集中的负载分布。

虽然负载均衡的本意是为了优化网络的运行,平均分配整个网络中的负载,达到尽可能高的网络资源的利用率,提高网络的性能。但是,作为一种网络管理行为,负载均衡为网络引入了负载信息交互的网络开销和负载均衡调整的网络管理开销,并且与之相关的,终端切换也会增加网络的开销。随着ESS中AP数量的增加或STA数量的增加,维护负载均衡所需的网络开销也会增大,对网络性能造成干扰和损失。

负载均衡也有负载调整失败的风险。切换有失败的可能,会造成孤立节点的存在。切换失败的终端不得不重新寻找AP进行接入,上层业务必然会中断。另外,切换是对稳定的网络环境的干扰,过于频繁的切换会造成网络性能的抖动,并且可能出现“乒乓效应”,使网络长时间处于动荡状态,不能稳定。

3 TD-SCDMA系统中的负载均衡技术

时分同步码分多址(TD-SCDMA)是一个时分同步系统,在通用移动通信系统陆地无线接入网(UTRAN)的无线资源控制(RRC)中,很重要的一部分内容就是负载均衡。在时分系统中,时间和码道资源是非常有限的,而且在时分系统中,大量的业务是电路交换业务,对QoS的要求都很高,所以,负载均衡在时分系统中非常重要。

时分系统中的负载均衡有多重粒度。除了可以控制用户终端(UE)的切换之外,甚至可以在每个时隙的基础上控制负载的分配。与无线局域网不同的是,时分中的某些负载均衡方法是以牺牲UE的业务为代价的,比如暂时中断分组数据业务,甚至强迫用户掉话,中断电路交换业务,强制UE释放系统资源。

一般而言,在时分系统中,可行的负载控制措施有:

- 强制使某些用户掉话。
- 在同一节点B(NodeB)的不同时隙间进行负载均衡。
- 下行快速负载功率控制:拒绝UE来的下行快速功率增加命令。
- 上行快速负载功率控制:减少上行快速功率控制使用的上行目标。
- 减少分组数据流量的吞吐量。
- 如果支持多载波,则切换到另外一个TD-SCDMA载波上。
- 切换到GSM或其他移动通信系统。
- 减少实施移动台的比特传输率,例如自适应码率(AMR)语音等。

• 降低基站发射功率,减小本基站覆盖,迫使本基站内的一些UE切换到其他小区。

其中,前4个措施可以在同一个基站内部执行,特点是快速,甚至可以在一个时隙内生效。为了保证那些不能忍受重发导致时延业务的QoS,可以提高对延时并不敏感的链路连接的误帧率要求,这样做尽管会增加分组数据业务的延时,但是却能维护诸如语音和可视电话等传统电路交换业务的质量。

可以看出,在时分双工的TD-SCDMA系统中,由于系统的粒度差异,导致负载均衡的方法也有很大的区别。但总体上,通常要经过负载信息收集、负载评估、任务跃迁3个部分。不同的是加上了码分多址(CDMA)和时分双工(TDD)的特点而已。

4 结束语

无线局域网中的负载均衡技术研究才刚刚开始。由于无线局域网的使用环境,负载均衡面临着比有线分布式系统中更加复杂的问题。好的负载均衡方法可以有效地提高网络的效率,提高资源的利用率,降低系统拥塞的可能性,提高系统抵抗负载变化的能力,提高系统的容忍度。无线局域网的负载均衡应该遵循以下原则:

- 准确地评估每个AP的负载状况,并尽快地进行负载信息的交互;负载信息的交互应该不影响业务质量。
- 高效的均衡算法,准确地选择执行负载均衡调整操作的执行者,并结合切换和接入两种手段;如何加快负载均衡调整的收敛速度也是均衡算法应该关注的重点。
- 结合快速切换的切换管理和控制,要避免多个终端在同一个AP处的切换竞争和拥塞。
- 降低负载均衡的副作用,应该降低网络开销,具备避免频繁切换、切换失败控制等机制。

5 参考文献

- [1] IEEE 802.11b. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications[S]. 1999.
- [2] GEIER J. 无线局域网[M]. 北京:人民邮电出版社, 2003.

收稿日期: 2005- 12- 27

作者简介



王彬,毕业于西安电子科技大学,硕士。中兴通讯股份有限公司中心研究院西安研究所高级工程师,主要从事IEEE 802.11和IEEE 802.16相关技术的研究和产品开发。