

doi:10.16055/j.issn.1672-058X.2015.0011.015

## 基于效用函数的异构网络负载均衡算法\*

郭 强, 车玉洁, 张晓萌, 朱若菡

(山东财经大学 管理科学与工程学院, 济南 250014)

**摘 要:**提出了一种基于效用函数的异构网络负载均衡算法.该算法通过网络参数的定时测量,得到各性能指标的效用函数,经过归一化和权重处理,选取重负载和轻负载的小区,并选取资源利用效率低的业务终端,通过终端的转移实现负载均衡.仿真分析表明,该算法能够有效提高网络吞吐量,降低网络阻塞率,并减少网络时延.

**关键词:**负载均衡;异构网络;效用函数;网络吞吐量;网络阻塞率

**中图分类号:**TN929.5      **文献标志码:**A      **文章编号:**1672-058X(2015)11-0061-06

未来的通信网络一定是多用接入网络并存的异构融合网络.异构无线网络的资源管理技术已成为热点研究领域,一个有效的无线资源管理机制能够充分利用有限的资源给用户提高质量的通信服务.负载均衡作为其中的关键技术之一,对于优化网络资源的使用,提高网络吞吐量和避免网络拥塞有着积极的意义<sup>[1]</sup>.

目前,对于异构网络负载均衡的研究很多,实现负载均衡的方式也有多种,可以根据整个网络系统的负载状况,通过业务流的转移来实现重负载和轻负载小区的网络负载均衡,也可以通过一定的接入规则或方案,为新业务提供最佳网络选择,避免接入负载过重小区,实现负载均衡<sup>[2]</sup>.文献[3]提出了基于业务类型和逗留时间的负载均衡算法,它根据吞吐量门限将小区的负载状态进行分级,根据不同负载状态接入不同的终端进而实现负载均衡.文献[4]提出了灰关联度负载均衡算法,将影响负载均衡的因素构建灰关联因子集并进行灰关联分析,得到网络的关联度,进而计算业务流在各网络之间的分配比例.这些算法从不同角度实现了负载均衡,并且这些算法中,大部分是根据网络负载状态进行判断是否进行负载均衡,对于负载状态即过重负载的门限值,没有具体的确定方式.在文献[5]中,提出了基于用户满意度的负载均衡算法,即当系统超过某一门限值时,把负载较重小区的终端移动到轻负载小区中,但是该方法采用的负指数函数作为终端收益函数,只是针对理想状态或特殊状态的研究,负指数分布只是爱尔朗分布的一种退化形式.在负指数分布的情况下,网络内的业务可以正常运行,但实际的业务流情况却要复杂的多.

针对上述问题,提出了基于效用函数的异构网络负载均衡算法,该算法通过网络参数的定时测量,得到各性能指标的效用函数(采用可以更好对现实数据进行拟合的 Erlang 分布模型),将其经过归一化和权重处理,选取重负载和轻负载小区,并选取资源利用效率低的业务终端,通过业务端从重负载小区到轻负载小区的转移实现负载均衡.

收稿日期:2015-02-26;修回日期:2015-06-02.

\* 基金项目:山东省自然科学基金(ZR2011FM022).

作者简介:郭强(1975-),男,山东人,博士,副教授,硕士生导师,从事无线通信网络研究.

## 1 基于效用函数的负载均衡算法

基于效用函数的负载均衡算法的核心思想:网络的负载状况和服务质量状况可以通过影响负载均衡性能指标的效用值反映出来,性能指标的效用值越大,则表明得到的服务质量越高,反之亦然.为了实现负载均衡并且提高用户终端的服务质量,算法将负载过重网络里效用较小的终端业务切换到重叠覆盖区域负载较轻的网络中.

### 1.1 性能指标及其效用函数

未来无线网络支持多种业务,其中包括语音、数据、视频流、网页浏览以及 FTP 文件传输等,这些业务可以分为实时业务(RT)和非实时业务(NRT)两类.为了分析简便,此处主要考虑两种典型的业务类型即实时语音业务和非实时数据业务.在评价网络的性能时,语音业务主要考察其呼叫阻塞率和掉话率,数据业务主要考察时延和丢包率,有效带宽是两种业务均考虑的因素.效用函数模型采用文献[6]的方法,首先假定两种业务的呼叫到达率分布服从均值为 $\lambda_{RT}, \lambda_{NRT}$ 的泊松分布,呼叫持续时间分别服从 $1/\mu_{RT}, 1/\mu_{NRT}$ 的指数分布,占用的数据长度服从 $\theta_{RT}, \theta_{NRT}$ 的指数分布.两种业务最多可以分别得到 $N_{RT}, N_{NRT}$ 个连接,根据 Erlang 分布模型,分别可以得到两种业务的呼叫阻塞率为

$$P_{RT/NRT} = \frac{(\lambda_{RT/NRT}/\mu_{RT/NRT})^{N_{RT/NRT}}/N_{RT/NRT}!}{\sum_{i=0}^{N_{RT/NRT}} (\lambda_{RT/NRT}/\mu_{RT/NRT})^i/i!} \quad (1)$$

另外,为了便于描述,将涉及的部分符号及含义标记于表 1.

表 1 文中使用的符号标记

参数符号	参数含义
$C_{RT/NRT}(i)$	网络 $i$ 为实时或非实时业务预留的带宽
$\rho_{RT/NRT}$	负载承受强度
$T_{RT/NRT}$	时延
$B_{avi}(i)$	网络 $i$ 的可用带宽
$B_{total}(i)$	网络 $i$ 的总带宽
$P_{RT/NRT\_max}$	业务能够容忍的最大阻塞率
$T_{RT/NRT\_avg}$	平均时延
$T_{RT/NRT}$	传输时延的一般值
$T_{RT/NRT\_max}$	最大时延

其中,负载承受程度及时延可以表示为

$$\rho_{RT/NRT}(i) = \frac{\lambda_{RT/NRT} * \theta_{RT/NRT}}{C_{RT/NRT}(i)} \quad (2)$$

$$T = \frac{\rho^{N+21}(N * \rho - N - 1) + \rho}{\lambda(1 - \rho^N)(1 - \rho)} \quad (3)$$

其中,式(3)中的各个参数分别指的是实时或非实时两种业务下的参数.

此处采用网络可用带宽、呼叫阻塞率和传输时延作为网络负载状况的衡量标准,下面对这 3 个参数的效用函数进行讨论.

网络  $i$  的可用带宽的效用函数可以表示为

$$U_B(i) = \frac{B_{\text{avi}}(i)}{B_{\text{total}}(i)} \quad (4)$$

从式(4)中可以看出,网络的剩余可用带宽越大,效用函数的值也越大.网络  $i$  的呼叫阻塞率的效用函数可以表示为

$$U_{\text{RT}}(i) = \frac{\exp(-P_{\text{RT}}(i))}{\exp(-P_{\text{RT}_{\text{max}}})} \quad (5)$$

式(5)为实时业务呼叫阻塞率的效用函数,非实时业务的同理.从式(5)可以看出,网络的呼叫阻塞率越小,效用函数的值越大.网络  $i$  的传输时延的效用函数可以表示为

$$U'_{\text{RT}}(i) = \exp\left(\frac{T_{\text{RT}_{\text{avg}}} - T_{\text{RT}}(i)}{1 + \sqrt{T_{\text{RT}_{\text{avg}}}/T_{\text{RT}_{\text{max}}}}}\right) \quad (6)$$

式(6)为实时业务传输时延的效用函数,非实时业务的同理.从式(6)可以看出,网络的传输时延越小,效用函数的值越大.

## 1.2 重负载和轻负载网络的选取

在进行重负载和轻负载网络选取时,需要根据不同网络的总效用值比较来确定其网络负载状况.对已剩余可用带宽、呼叫阻塞率以及时延各个参数进行归一化处理,让其取值在(0,1)之间.

对于  $U_B$ ,由于该效用值已经在规定范围内,所以不必再进行归一化处理.对于  $U_{\text{RT}}$ 和  $U'_{\text{RT}}$ ,采用如下归一化方法进行处理:

$$\bar{U} = \frac{U_i - U_{\text{min}}}{U_{\text{max}} - U_{\text{min}}} \quad (7)$$

其中, $U$ 代表  $U_{\text{RT/NRT}}$ 或者  $U'_{\text{RT/NRT}}$ , $\bar{U}$ 为归一化后的效用函数值, $U_{\text{max}}$ 和  $U_{\text{min}}$ 分别代表对应参数的最大值和最小值.因此归一化之后,对于实时业务,3种效用函数值分别为 $\bar{U}_B$ , $\bar{U}_{\text{RT}}$ 和 $\bar{U}'_{\text{RT}}$ .对于非实时业务,3种效用函数值分别为 $\bar{U}_B$ , $\bar{U}_{\text{NRT}}$ 和 $\bar{U}'_{\text{NRT}}$ .由于不同的业务类型考察不同的参数,实时业务更看重呼叫阻塞率,非实时业务更看重时延.因此,将实时业务和非实时业务分开进行讨论.

对于实时业务,效用  $U_1 = k_1 \bar{U}_B + k_2 \bar{U}_{\text{RT}} + k_3 \bar{U}'_{\text{RT}}$ ,其中  $k_1 + k_2 + k_3 = 1$ , $k_2$  占较大比重.对于非实时业务,效用  $U_2 = k^1 \bar{U}_B + k^2 \bar{U}_{\text{NRT}} + k^3 \bar{U}'_{\text{NRT}}$ ,其中  $k^1 + k^2 + k^3 = 1$ , $k^3$  占较大比重.因此,网络  $i$  的总效用为

$$U(i) = U_1(i) + U_2(i) \quad (8)$$

为了使负载均衡快速收敛,算法在迭代过程中需找出全网负载最重的接入网,选择其合适的用户终端切入到其他网络中.而网络效用能够体现负载状况,因此选择效用最小的网络为重负载网络即待减负网络,将其记为  $H$ ,则

$$H = \operatorname{argmin}\{U(i)\} \quad (9)$$

同理,选择效用值最大的网络为轻负载网络即待接入网络,将其记为  $L$ ,则

$$L = \operatorname{argmax}\{U(i)\} \quad (10)$$

## 1.3 业务流的选取

在找出全网负载最重的网络后,需要从此网络中选取合适的业务流切换到其他接入网中.在选取业务流时,要考虑两方面的因素:当前业务流的服务质量状况;当前业务流的资源利用效率,即倾向于选择服务质量体验较低或者信道状况差的业务流切换到其他网络以得到更好的服务质量,并停止在重负载网络的资源占用.选择业务流直到重负载网络的效用值和轻负载网络的效用值尽可能接近为止.

## 1.4 算法流程

此处提出一种基于效用函数的异构网络均衡算法的流程图如图 1 所示.

步骤 1:网络管理中心保持实时监测各个网络的负载状况,当达到预定时间时,触发基于效用的负载均衡算法.设定预定时间触发机制可以在负载过重的情况发生前合理配置资源,一定程度减少过重负载情况发生.

步骤 2:管理中心获取网络参数值,包括实时和非实时业务的数量等.

步骤 3:计算剩余可用带宽、呼叫阻塞率和时延的效用函数,并进行归一化处理.

步骤 4:选择负载最重和负载最轻的网络分别作为待减负网络和待接入网络.

步骤 5:在重负载网络中选取合适的业务流进行切换.

从算法的流程看出,通过效用函数的设计,网络效用和业务流的收益可以量化衡量.通过网络和业务流的选择及切换,网络的效用和业务流的 QOS 收益逐渐实现均衡.付出的代价只是由于算法的运行,需要网络和管理中心实际执行,并交换相应的信息和网络选择策略.可以看出,此处提出的算法可以以较小的代价换取整个网络性能的提升.

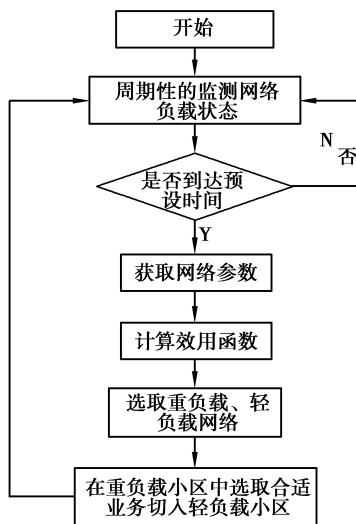


图 1 算法流程图

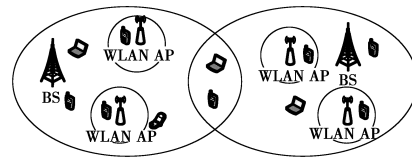


图 2 异构网络场景图

## 2 仿 真

首先介绍仿真的评估模型与参数设置,然后对实验结果进行分析.为了对基于效用的负载均衡算法的性能进行全面的分析和评价,采用随机游走模型描述异构网内用户终端的移动规律.异构网络的仿真场景如图 2 所示.图 2 所示的异构网络场景图由 2 个 UMTS 网络及 4 个 WLAN 网络形成交织重叠覆盖,大量用户在重叠异构覆盖范围随意行走.设定系统中存在 60 个用户终端,每个用户只使用一种业务,每种业务各 20 个用户.业务的分组到达服从泊松分布,用户终端同一时刻只能接入一种网络,用户初始接入按照接入信号强度准则进行,即根据用户终端接收到的接入网络的信号强度,动态选择信号强度最大的网络接入.用户终端在异构网中,依据此处提出的基于效用函数的算法选择不同的网络实现负载均衡.表 2 给出了仿真参数及其取值.

表 2 仿真参数及数值

参 数	数 值	
	UMTS	WLAN
RSSI/dBm	-20~-125	-45~-95

续表

参 数	数 值	
$R$ (小区半径)/m	1 000	200
$B$ /Mbps	0~2.5	0~11
呼叫到达率 $\lambda$	实时	非实时
呼叫持续时间 $1/\mu$	1~5	1~10
呼叫持续时间 $1/\mu$	0~100	0~500
负载均衡触发预定 时间/s	100	

为了验证所提出的基于效用函数的负载均衡算法(ULBA)的性能,采用文献 7 提出的基于 QOS 感知的算法(QLBA)作为对比算法,从实时业务阻塞率、非实时业务时延以及全网吞吐等方面进行比较.图 3-5 为仿真结果.其中图 3 比较了 ULBA 和 QLBA 两种算法下的实时业务的阻塞率和业务到达率的关系.可以看出,随着呼叫到达率的增加,由于网络资源有限,实时业务的阻塞率有上升趋势.但 QLBA 较新算法 ULBA 下的阻塞状况更严重.由此看出,新算法 ULBA 可以及时调节网络间的负载均衡,给用户带来更好的服务质量.

图 4 给出了 QLBA 和 ULBA 两种不同的算法下非实时业务传输时延的对比结果.可以明显看出,两种方法下非实时业务的传输时延均随着用户呼叫到达率的提高而增加,但是新算法 ULBA 下的时延相对 QLBA 来说更短,让终端用户等待的时间也更短,服务质量更高,能够说明新算法 ULBA 在缩短时延方面有更好的性能.图 5 给出了 QLBA 和 ULBA 两种算法下网络吞吐量的情况,可以看出,当业务流速率较低时,两种算法差异不大,但随着业务流平均速率不断增加时,采用 ULBA 算法使网络有更大的吞吐量.

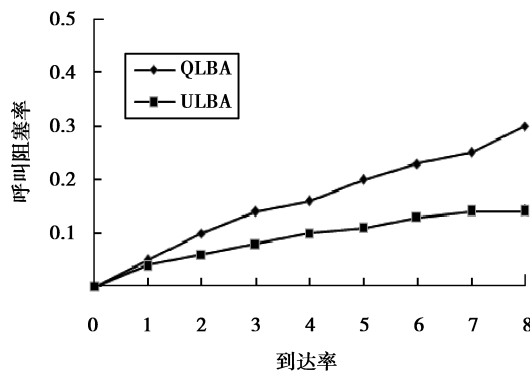


图 3 实时业务阻塞率

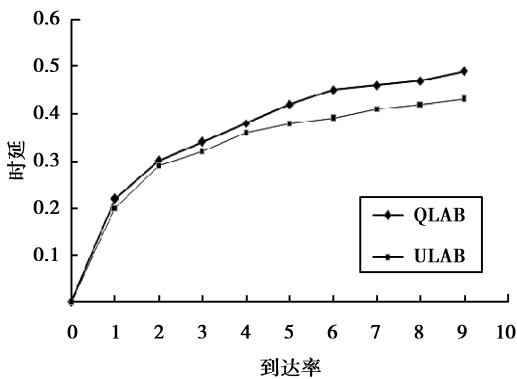


图 4 非实时业务传输时延

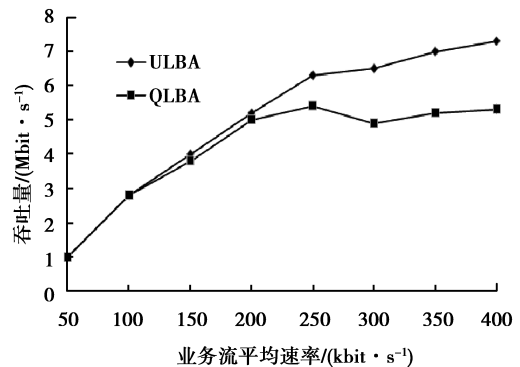


图 5 全网吞吐量

### 3 总 结

提出了一种基于效用函数的异构网络均衡算法,该算法通过网络参数的监测测量,并通过预定时间触发算法,首先获取各网络参数的相关信息,得到一些性能指标的效用函数(采用可以更好地对现实数据进行拟合的 Erlang 分布模型),经过归一化和权重处理,选取重负载和轻负载的小区,并选取服务质量状况差或资源利用效率低的业务终端,通过终端的转移实现负载均衡.仿真分析表明,该算法能够有效提高网络吞吐量,降低了网络阻塞率,并减少网络时延.

#### 参考文献:

- [1] CIUBOTARU B, MUNTEAN G M. A Quality-oriented Handover Algorithm for Multimedia Content Delivery to Mobile Users [J]. IEEE Transactions on Broadcasting, 2009, 55(2): 437-450
- [2] ZAHARAN A H, LIANG B. A Generic Framework for Mobility Modeling and Performance Analysis in Next-generation Heterogeneous Wireless Networks [J]. IEEE Communications Magazine, 2007, 45(9): 92-99
- [3] 罗俊辉,白光伟,沈航,等.异构无线网络终端服务感知的动态负载均衡机制[J].计算机科学,2014,41(6):37-42
- [4] 陈晓玉.异构无线网络的负载均衡算法研究[D].南京:南京邮电大学,2013
- [5] 王华东,侯燕,王凤春.一种适用于 Ad hoc 网络的基于概率负载均衡算法[J].计算机应用研究,2014,31(3):872-875
- [6] 焦毅,易克初,马懋德,等.基于服务质量感知的异构无线网络负载均衡算法[J].吉林大学学报,2013,43(3):794-800
- [7] 周炯槃,张琳,望育梅,等.通信网理论基础[M].北京:人民邮电出版社,2009
- [8] 刘琪,袁坚,山秀明,等.3G/WLAN 网络中基于终端移动与业务认知的动态负载均衡机制[J].计算机学报,2010,33(9):1569-1579
- [9] 石文孝,张阁,王继红,等.基于网格的异构无线网络负载均衡算法[J].吉林大学学报,2013,43(3):788-793
- [10] ZUKRI M, JOUABA B, ZEGHLACHE D. A Review on Mobile Management and Vertical Handover Solutions over Heterogeneous Wireless Networks [J]. Computer Communications, 2012, 35(17): 2055-2068
- [11] LEE D, HSUEH Y. Bandwidth-reservation Scheme Based on Road Information for Next-generation Cellular Networks [J]. IEEE Transactions on Vehicular Technology, 2004, 53(1): 243-252
- [12] 孙建民.无线传感器网络[M].北京:清华大学出版社,2006
- [13] PIAMRAT K, KSENTINI A, BONNIN J, et al. Radio Resource Management in Emerging Heterogeneous Wireless Networks [J]. Computer Communications, 2011, 34(9): 578-585
- [14] 彭燕.凹印机 P2D 算法实现[J].包装工程:工程版,2015(17):131-133;138

## Load Balancing Algorithm for Heterogeneous Wireless Networks Based on Utility Function

**GUO Qiang, CHE Yu-jie<sup>\*</sup>, ZHANG Xiao-meng, ZHU Ruo-han**

(School of management Science and Engineering, Shandong University of Finance and  
Economics, Jinan 250014, China)

**Abstract:** Load balancing algorithm for heterogeneous wireless networks based on utility function is proposed in this paper. Through timing measurements of network parameters, the algorithm firstly gets the utility function of each performance index, and secondly deals with the indexes by normalization and weighting process, then selects the busiest RAN and lightest RAN, and also selects the low QOS utility terminal, finally achieves load balance by transferring terminal. Simulation results show that the proposed algorithm can effectively improve network throughput and reduce network congestion rate and time-delay.

**Key words:** load balance; heterogeneous network; utility function; network throughput; network congestion rate