International Journal of Applied Mathematics and Soft Computing

Volume 2, Issue 2, August 2016

Publisher: HongKong New Century Cultural Publishing House
Address: Unit A1, 7/F, Cheuk Nang Plaza, 250 Hennessy Road, Wanchai, Hong Kong
Tel: 00852-28150191
Fax: 00852-25445670

Contents

Load Balancing Algorithm based on QoS Awareness in Wireless Networks

Jianjun WU
Hunan City University, Yiyang Hunan 413000, CHINA

Abstract: In order to reduce business latency and packet loss rate, to improve the throughput of integration of heterogeneous wireless network, to achieve load diversification and to improve end quality of service, while there are many problems on the dropout rate of the traditional load balancing algorithm in processing throughput, delay and business, therefore, gateway load balancing algorithm is proposed in this paper. All terminals in the access network can reflect the effectiveness of the average network load level of the network, this algorithm will gain weight of the load in the network of small business terminal to switch to the network load which is light. First, it defines the heterogeneous networks and network terminal payoff function utility function which are used to characterize the experience and network QOS terminal load situation, and then presents the specific processes of the gateway load balancing algorithm. Finally, switching decisional load balancing algorithm, proposed by Yan X, etc. is compared with simulation experiments and it shows that: the proposed gateway load balancing algorithm has strong robustness to achieve network load balancing and to achieve a balanced use of network resource.

Keywords: Efficiency; Loading differences; Real-time business; Switching decision

1. Introduction

Load balancing is an important way, which makes full use of heterogeneous wireless network resources. Through the load balancing among the heterogeneous networks, the high probability of network load can be reduced, the overall utilization of network resources is improved and the blocking probability is reduced to provide users a better QOS, so the load balancing between the networks is also an important aspect of considering the access algorithm selected [1-3].

The user can only access a kind of network at the same time with the traditional hard load balancing service, which can not satisfy differentiation of the user service, resulting in partially utilized heterogeneous network resources and higher traffic blocking ratio. Literature [4] proposes the soft load balancing algorithm, which dynamically changes the best ratio of IP flow; when the network load is heavy, each network sub-flow rate of accessing users is consistently improve, while when the network load is light, there is a differences between the needs of users business rate and wireless resources [5-7].

2. Proposed Scheme

The core idea of algorithm is: the QOS income of terminal end can reflects the current level of network quality of service received, the greater the benefits is, the better the resulting quality of service is, and vice versa; the average QOS benefit of all terminals in a accessing network (hereinafter defined as a network utility) can reflects the load level of the network, the larger the average gain of the terminal is, the load on the network is lighter, and vice versa. In order to achieve diversification and to increase the service quality of load terminal, the algorithm will switch small businesses with gain heavy load in the network to a lighter terminal network load.

In order to carry out the terminal to select the RAN and its access effectively and dynamically, the multimode terminal will experience unified management from the network side, which will be completed by the network side management entity (Network side manager, NSM). The terminal will interact with NSM through the terminal-side management entity (Terminal-side manager, TSM), dynamically achieving refactoring of switching/accessing. The interaction between NSM and TSM is completed through management control channel (Management and control channel, MCC). NSM is deployed in the core network, and is shared by a plurality of RAN. RAN will convey each context information to NSM, then NSM transfer each context information of RAN to the terminal for the decisions through the MCC downlink transmission. TSM of each terminal sends context information of the terminal through uplink transmission of the MCC to NSM. Based on context information of RAN and terminal, NSM uses the appropriate network selection algorithm to develop strategies and policies issued under the various terminals. Terminal then chooses according to their needs and network reconfiguration decisions and configures to access the appropriate RAN. The paper will assume the terminal in the network / inter-cell handover fast enough, and thus, the load balancing process, due to switching delay caused by the upper risk of business disruption can be ignored.

2.1. QoS revenue and RAN utility

Different types of wireless network services have different QOS requirements. Based on the characteristics of the various services, wireless services can be classified into three types of basic services, as shown in Table 1.

Table1. Wireless network traffic types and their QOS requirements

	Rate of change	The minimum rate	Maximum rate	Time delay requirements	A typical business
Real-time business	constant	$r_{\min} = r_{\max}$	$r_{\min} = r_{\max}$	$d_{\max}$	VoIP
	variable	$r_{\min}$	$r_{\max}$	$d_{\max}$	MPEG
The real-time business	variable	$r_{\min}$	$r_{\max}$	φ	HTTP

For real-time services, at the premise to meet the minimum bandwidth, using an average delay d_{ij} to measure user gains, the smaller the time delay is , the higher the gain is; for non-real-time services to the user, using average speed r of the user gains to measure, the larger the rate is, the higher the income is;

$$\frac{d_{ij} - d_{ij}^{e}}{d_{ij}^{\max} - d_{ij}^{r}}$$

If delay for real-time services the normalized, wherein d_{ij}^{e} denotes the expectancy of average delay of real-time services:

$$\frac{\beta_{ij} - \beta_{ij}^{e}}{\beta_{ij}^{\max} - \beta_{ij}^{r}}$$

As the normalization about Non-real-time services' rate, it helps to ensure a minimum rate of non-real-time services; β_{ij} and β_{ij} is constant parameters, which determines the steepness of the curve of the function, the larger the value is, the steeper the curve changes, the higher the sensitivity to the end quality of service is. Formula (1) as defined in revenue function reflects QOS-awareness of terminal, the function maps a plurality of QOS parameters with reasonable perception or experience for the user to QOS level, gives a measurement of the QOS of different users by using uniform quantization levels standards.

To characterize the load level of the accessing network, the wireless access network defines the utility of all the terminals connected to the network average of QOS benefit. Suppose at a time, a terminal can only access a RAN, then the gain RAN $J \in I$ can be expressed as

$$R_{ij} = \begin{cases} \dfrac{\sum_{j \in i} \alpha_{ij} n_{ij}}{\sum_{J \in I} \alpha_{ij}}, & \sum_{I \in J} \partial_{ij} \neq 0 \\ 1, & \sum_{i \in j} \partial_{ij} = 0 \end{cases} \quad (1)$$

Wherein, $\partial_{ij} = \begin{cases} 1 \text{ terminal } i \text{ int} o\ RAN_j \\ 0 \quad other \end{cases}$

Obviously, the heavier network load will result in lower average QOS benefit of terminal, otherwise the terminal average QOS gains will be higher. Therefore, the average QOS benefit of the terminal, namely the network load of the network utility can reflects the situation. Network utility is higher, indicating that the network load is lighter, otherwise it indicates the network load is heavier.

2.2. Load balancing algorithm design

1) Web-based utility accessing network selection to be burdens

In order to make load balancing rapidly converge, algorithms in each iteration find the heaviest load of the whole network accessing network, and select one of the appropriate terminal switch to other accessing networks.

According to the definition and analysis of the effectiveness of the network utility in the section of 3.1, it can effectively reflect the access network load conditions. Therefore, the algorithm selects the lowest utility network accessing network as a network to be burdens. If pending burdens network denoted $J \in I$, then:

$$r^{n} = \arg\min_{j \in J} \left\{ \bar{m}_{j} \right\}$$
$$r^{i} = \arg\min_{i \in I} \left\{ \bar{m}_{i} \right\} \quad (2)$$

2) The weighted selection of subjecting *to switch terminal about QoS benefit to be based on*

In order to make load balancing rapidly converge, algorithms in each iteration find the heaviest load of the whole network accessing network, and select one of the appropriate terminal switch to other accessing networks. When the accessing network subject to reducing the burdens selects to switch's terminal, it not only needs to consider the quality of service experience of terminal at the corresponding to the accessing network, namely it should choose a lower the QOS experience in order to the user expects to switch to other network medium to get improved; and it needs to consider the users' radio resource utilization efficiency to be burdens in the accessing network, which should tend to choose the channel conditions of poor terminal to switch to other networks, in order to terminate their treatment on resources inefficient usage in subjecting to reducing the burdens in the accessing network channel. Considering two aspects as defined in the terminal i in the weighted QOS benefit function of the RAN j is as follows:

$$J_{ij}^{*} = a_{ij} \times J_{ij}$$
$$J_{ij}^{*} = (a_{ij} + J_{ij}) \quad (3)$$

Wherein the weighting factor a_{ij} reflects resource utilization efficiency of the terminal i in the radio accessing network j, which is defined as

$$a_{ij} = \frac{J_{ij}(e_{ij})}{J_{ij}^{\max}}$$
$$e_j^i = \sum_{g}^{a_{ij}} \beta u_i^j \quad (4)$$

$J_{ij}(e_{ij})$ is the actual rate that each unit bandwidth obtained for the accessing terminal i to the accessing network j ; it indicates channel condition between the terminal i and accessing network j ; $R_{i,j}^{\max}$ presents the maximum rate can obtain per unit of bandwidth from accessing terminal i to the accessing network j theoretically. Thus, the greater a_{ij} is, indicating that the channel conditions is better between terminals i and accessing network j the terminal i has a good utilization efficiency in the accessing network channel .

3) The selection of the accessing network subject to be negative growth

When the terminal to be switched e_j^i selects e^i, the intending accessing network, for the purpose of switching accessing networks, on the one hand, it needs to identify whether the terminal e_j^i is within the coverage area of the accessing network e^i , and the remaining capacity of e^i can meet the minimum rate requirements of e_j^i. That the intending accessing network e^i shall meet

$$\begin{cases} e_{i^n}(r_n^i, e^i) \succ 0 \\ e_{ij}^{\max} + \sum_{j \in i} \partial_{ij}^{\min} \le r_{ij}^{\max} \end{cases} \quad (5)$$

Among which, $r_{ij}^{\max}$ represents the maximum network capacity of accessing networks e^i.

4) QoS-awareness load balancing algorithm

Based on the above analysis, the designation of gateway load balancing algorithm process is as follows:

Initialization: Let J represents algorithm converged network within the scope of all the heterogeneous collection of RAN, I represents the set of all terminals in the network; / / Algorithm carries through load balancing to a certain geographical area of heterogeneous converged network.

Step 1: Select (3) to satisfy the access network e^i as a network to be the network subject to reducing the burdens; e_j^i as a collection of all terminals e^i connected, optionally switch the terminal to be set $e_j^i = \overset{-i}{e}_j$.

/ / selects the heaviest accessing network, ready to be moved to another part of the load in the accessing network.

Step 2: In $\overset{-i}{e}_j$, select the terminal $e^i = \arg\min_{i \in m}\{e_{ij}^{*}\}$ as to be switching terminals.

/ / selects the terminal with a lower service quality of experience and resource utilization, ready to be switched to another accessing network.

Step 3: among other accessing network options except e^i, select e_j^i to meet (7) formula and (8) as accessing network to be negative growth.

/ / selects an accessing network terminal with a lighter load to be switched as the intending accessing network.

Step 4: switch the terminal to be switched e^i to the accessing network to be negative growth e_j^i, if

$$e_j^i \succ e_{i^h, j^h}$$

The terminal e^i switch to e_j^i, then return to step 1; otherwise terminal e^i going back to e_j^i, while the optional terminal to be switch can is updated to $\overset{-i}{e}_j = e_j^i - \{e^i\}$ and perform step 5.

/ / If the accessing network can be increased to improve the negative terminal of the quality of service to be switched, the adjustment is successful and re-select the accessing network of subjecting to reducing the burdens to carry out a new round of adjustment; otherwise the terminal to be switched from the optional subject can be removed from the terminal concentrator.

It can be seen that network utility and terminal benefits can achieve quantitative balance from the algorithm through a unified QOS revenue function designation. By choosing and switching network and terminal QOS benefit between the network utility and terminals tends to equilibrium. The cost is based on algorithm designation, need to run on the network and terminal management entity execution algorithm, and the corresponding QOS information and interactive network selection strategy. Network side and terminal side executing algorithm, network management entity, can be naturally and easily completed by heterogeneous network convergence in the framework of NSM and TSM, interaction of information and strategies can be completed by the contextual messages of the MCC. In the current level of technology, NSM and TSM itself has a powerful computing capabilities, and load balancing does not require strict real-time and interactive information and data of selection strategies for QOS is a small amount, therefore the proposed algorithm can obtain improved network and terminal performance significantly at little cost.

3. Experimental Results

By using OMNET + + 4.0, it can evaluate its performance of load balancing algorithm based on QOS-awareness. Supposing that, there are two types of 802.11 and 802.20 RAN in the system, each type of RAN has 2. Simulation only consider the upstream traffic, 802.20 RAN adopts the TDD mode of single carrier , each TDD frame in the uplink and downlink frames has each half; 802.23 RAN uses DCF mode. There are 30 supposed user (terminal), each user uses only one business, each has 10 business users. Setting the density of user in 802.20 RAN, the user from the SNR to the RAN randomly changed. Users can access a RAN at the same moment. Service packets arrive to Poisson distribution. The minimum rate of Variable rate services reached 1/2 of average rate. Users initially access network, whose selections are based on SINR (Signal-to-interference plus noise ratio, SINR) criterion, that is, channel conditions between the user and the various RAN, dynamically select the highest RAN of SINR to access, while according to network selection principle of 802.20/802.23 converged network users to access, business users access to 802.20 networks real-timely at a constant rate. The associated parameter settings used by network simulation are shown in Table 2.

Table 2. Experimental parameters

Parameter names	The parameter value	Parameter names	The parameter value
The 802.20 frame length/ms	1	Packet length/bit	1000
The 802.20disabled when the time/ms	0.2	w_{xy} /ms	0
802.20 Frame when the number of disabled/ms	5000	w_{xy} /ms	100
The 802.23disabled when the time/ms	20	η_x	0.53
The SIFS length/ms	10	η_y	0.02
The DIFS length/ms	53		

In order to verify the performance of the proposed gateway load balancing algorithm, by using handover decision load balancing algorithm (Handoff decision load-balancing algorithm, HDLBA) proposed by Yan X, etc as a compared algorithm, the packet loss rate, packet delay and throughput performance were compared. Figure 1 - Figure 4 shows the simulation results of the comparison.

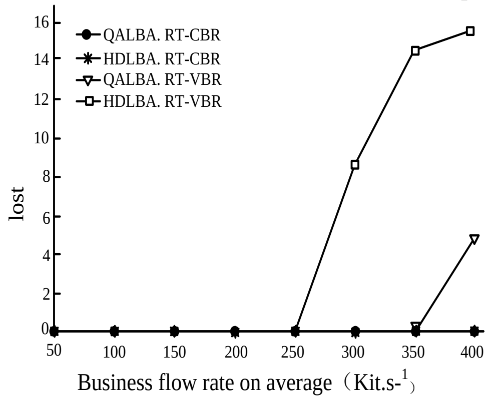

Figure 1. Real-time service of packet loss rate

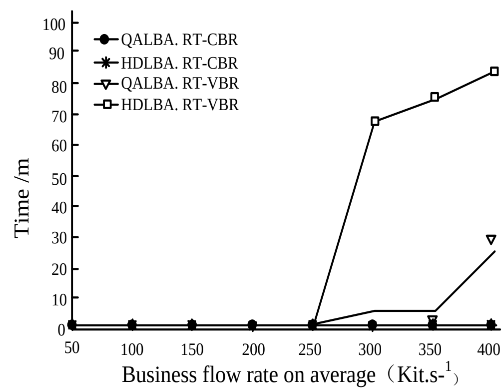

Figure 2. The average packet delay of time business

Figure 1 and Figure 2 show that the changing situation of real time constant bit rate and real time constant variable bit rate, average packet loss rate and packet delay variation when the average rate of each connection from 50Kbit /s increased to 400Kbit /s. As can be seen, the proposed gateway load balancing algorithm significantly reduces the real-time variable rate, packet loss rate and packet delay; while the average delay and packet loss rate of real-time business a constant rate in the two algorithms is basically to keep unchanged. It is because the priorities safeguard mechanism of 802.16 networks for real-time business with a constant rate can guarantee QOS service of real-time at a constant rate, while the two algorithms, which can ensure a constant rate of real business users, always stay in the 802.16 network, and get a guaranteed quality of service.

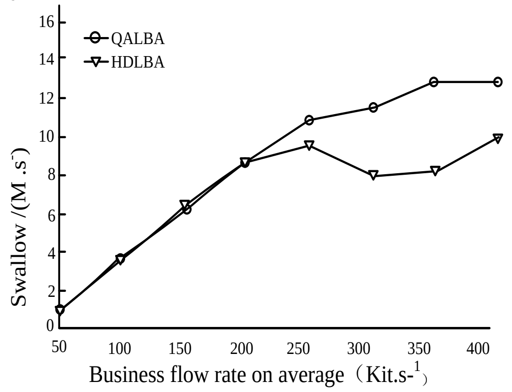

Figure 3. The throughput of whole network

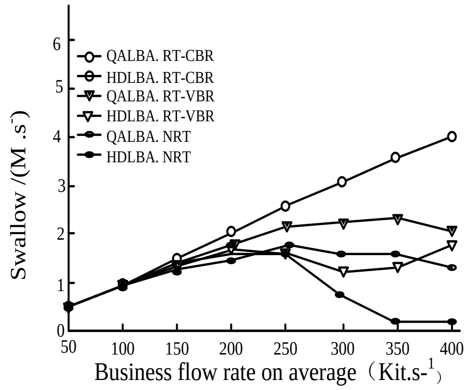

Figure 4. The throughput of various business

Figure 3 shows as the load increases, the proposed gateway load balancing algorithm can significantly improve the throughput of the whole network. Figure 4 shows the throughput of the two algorithms of different business changes. For real-time and non-real-time variable rate business operations (Non Real Time, NRT), QALBA algorithm's performance is significantly better than HDLBA algorithm; and this algorithm can always meet the real needs for traffic throughput at a constant rate. With the increase of user's average traffic rates, the network throughput of the users is gradually increased, because the QALBA algorithm can split the IP flow of the users service into two sub-streams, which can be inserted two networks at the same time, when the load of the two nets are high. The load balancing between networks is achieved and the network resources are rational allocated when the load is relatively light, so the throughput of HDLBA algorithm is lower than the throughput of other four algorithms. The load balancing of QALBA algorithm is the strongest, HDLBA algorithm the second and HDLBA algorithm the worst. Simulation results show that: QALBA algorithm can balance the network load, and compared to the traditional MLB algorithm and DLBD algorithm, the average blocking rate of packet service and the throughput performance have been improved significantly, and the robustness of algorithm is really strong, which can achieve network load balancing to achieve a balanced use of network resources.

5. Conclusion

For the integration of heterogeneous wireless network communication scenario, the network load balancing algorithm was proposed based on the QOS, which is fallen into switchable load balancing algorithm. Algorithm is based on QOS gains and network utility of a generalized feature for wireless business users, and it is able to characterize the quality of service terminal network experience and network load conditions, and it is universally applicable for a variety of heterogeneous network. Among them, the universality of network utility to heterogeneous network makes heterogeneous network load comparability comparable, able to achieve switching between heterogeneous network load balancing. Simulation results show that the proposed gateway load balancing algorithm can improve integration of heterogeneous network throughput, reduce business latency and packet loss rate, with strong robustness to achieve network load balancing and to achieve a balanced use of network resources.

References

[1] Kasman Suhairi, Ford Lumban Gaol, The Measurement of Optimization Performance of Managed Service Division with ITIL Framework using Statistical Process Control. Journal of Networks, Vol 8, No 3 (2013), 518-529

[2] Guang Yan, Zhu Yue-Fei, Gu Chun-Xiang, Fei Jin-long, He Xin-Zheng, A Framework for Automated Security Proof and its Application to OAEP. Journal of Networks, Vol 8, No 3 (2013), 552-558

[3] Muhammad J. Mirza, Nadeem Anjum.Association of Moving Objects Across Visual Sensor Networks.Journal of Multimedia, Vol 7, No 1 (2012), 2-8

[4] Muhammad J. Mirza, Nadeem Anjum.Association of Moving Objects across Visual Sensor Networks. Journal of Multimedia, Vol 7, No 1 (2012), 2-8

[5] Zhao Liangduan, Zhiyong Yuan, Xiangyun Liao, Weixin Si, Jianhui Zhao.3D Tracking and Positioning of Surgical Instruments in Virtual Surgery Simulation. Journal of Multimedia, Vol 6, No 6 (2011), 502-509

[6] S. H. Tang, M. C. Chen, Y. S. Sun, et al. A spectral efficient and fair user-centric spectrum allocation approach for downlink transmissions.IEEE., Globecom.,2011:1-6

[7] C. Xu, J.H. Lu, Y. Zhen. "An Experiment and analysis for a Boiler Combustion Optimization on Efficiency and NOx Emissions," Boiler Technology, Vol.37, No5,pp. 69-74, Oct. 2006.