# Comparative Performance Analysis of Dynamic Load Distribution Algorithms in a Multi-AP Wireless Network

Eng Hwee Ong, Jamil Y. Khan and Kaushik Mahata
School of Electrical Engineering and Computer Science
University of Newcastle, Australia, NSW 2308
Email: enghwee.ong@studentmail.newcastle.edu.au, {jamil.khan, kaushik.mahata}@newcastle.edu.au

***Abstract*—With the increased deployment of WLAN access points (AP), the importance of IEEE 802.11x standard network is also increasing. Mass deployment of WLANs offers the advantage of increased QoS support for multimedia traffic by using an intelligent vertical handover technique combined with an advanced admission control algorithm. In this paper we analyze the performance of three different dynamic load distribution algorithms which can be used in a multi-AP based heterogeneous wireless network. Two of the algorithms are based on prediction technique and the third one is a reactive algorithm. In all cases, QoS-based vertical handover is initiated only after QoS degradation of a station is detected. All three algorithms are simulated using OPNET to compare their performances.**

## I. Introduction

IEEE 802.11 WLAN is one of the de-facto wireless access networks offering broadband connectivity, thanks to its pervasive deployments over many diverse environments. The forthcoming 802.11n standard will further accentuate its benefits for high-speed ubiquitous broadband wireless access. However, delivering QoS demanding applications such as voice over WLAN (VoWLAN) are very challenging, particularly in the context of a hotspot configuration where multi-AP are physically co-located.

In general, network operators are motivated to maximize their revenue by maintaining a high system utilization while the end-users demand good QoS. It is known that QoS would inevitably deteriorate when network is driven beyond its capacity limits. Hotspots are typically deployed under such circumstances to cope with heightened traffic demands. However, the composite network capacity will not scale with the increasing number of APs if stations select AP based on received signal strength only, without QoS considerations such as load control or an appropriate network control mechanism such as admission control. This problem is further complicated by the typical non-uniform load distribution across APs in public hotspots such as convention centers and airports where users tend to correlate temporally and spatially. Consequently, these cause sporadic congestions in APs with the strongest signal strength. Therefore, load and/or admission control must be incorporated in such multi-AP hotspots so that diversity could be exploited to harness composite network capacity and QoS improvements.

Load control is typically concerned with load distribution to improve network QoS performance by transferring stations from heavily to lightly loaded networks. This allows stations to take advantage of the spare network capacity which would otherwise be left unused. Admission control is also critical for provisioning of QoS by regulating input traffic and preventing overloading of network. It works by conducting an assessment to check whether a new flow can be admitted without compromising the QoS requirements of existing flows. Hence, admission control policy dictates the provisioning of either guaranteed or predictive QoS. In fact, admission control and load control are often not dissociable. The main reason is that both rely on the knowledge of load metric in order to make their decisions. Henceforth, we treat both load and admission control interchangeably in the context of this paper.

Load distribution algorithms can be broadly classified as static, dynamic or adaptive. The main difference between static and dynamic load distribution algorithms is that the latter utilizes system state information, which enables exploitation of short-term fluctuations, to improve the quality of their decisions. Dynamic load distribution algorithms can be further categorized as load balancing or QoS balancing algorithms. Both algorithms have the same primary function of avoiding under-utilized networks when distributing load. The subtle difference is the former attempts to equalize load while the latter attempts to equalize QoS across networks in order to improve QoS for all flows. Adaptive load distribution algorithms are an extension of dynamic load distribution algorithms with the additional capability to adapt their parameters or policies dynamically in response to the varying system state.

Although various load distribution algorithms for WLAN have been investigated in literature, there is a lack of performance comparison between different algorithms. In this paper, we present a comparison of three dynamic load distribution algorithms, viz. one belonging to the class of load balancing algorithm and the other two belonging to the class of QoS balancing algorithm for infrastructure-based WLAN with DCF access mechanism. The remainder of the paper is organized as follows. Section II discusses related work. Section III describes the different dynamic load distribution algorithms. Section IV illustrates comparative performance evaluations. Section V presents conclusions and future work.

978-1-4244-4859-3/09/$25.00 ©2009

## II. Related Work

A key issue in designing any load or admission control algorithms is to identify a suitable load metric to estimate the available network capacity accurately. Bianchi *et al.* [1] first introduced the notion of 'packet level' load metric and showed that load balancing of WLAN can be improved.

Derivation of packet level load metrics could be categorized in two main threads, viz. model-based and measurement-based. In model-based approach, packet level load metrics are obtained by analyzing the WLAN DCF using the two-dimensional Markov chain model with or without theoretical queueing models. Zhai *et al.* [2] integrated Bianchi's model [3] with M/M/1/K and M/G/1/K queueing models to give non-saturation throughput, packet delay and loss bounds. The authors also showed that although M/G/1/K model provides better accuracy than M/M/1/K model in general, they do not exhibit significant difference in the non-saturation region. Malone *et al.* [4] extended Bianchi's model to non-saturation conditions by incorporating post-backoff states under bufferless network assumption. The authors also considered stations with different arrival rates but with same packet lengths. In measurement-based approach, packet level load metrics are obtained by either direct measurements or estimations from the system itself. Velayos *et al.* utilized throughput per AP to reflect the load of a network. Ong *et al.* [5] employed packet delay per AP to capture both network and channel variations which are indicative of the network load. Above all, channel utilization estimation first proposed by Garg *et al.* [6] gave the best representation of the effective network load.

The level of centralization also plays a crucial role in dynamic load distribution algorithms. Balachandran *et al.* [7] presented an adaptive load balancing solution where a centralized admission control server contains load information of all access points. Velayos *et al.* [8] proposed a decentralized load balancing scheme where access points are then classified based on their throughput in one of the three states, viz. underloaded, overloaded or balanced. It is known that both centralized and decentralized architectures have their pros and cons. Recently, a terminal-oriented network-assisted (TONA) handover architecture, which is a compromise between centralized and decentralized ones, is proposed in [5].

## III. Dynamic Load Distribution Algorithms

The comparison of the three dynamic load distribution algorithms is summarized in Table. I. Since these algorithms span across different levels of centralization, their performances are investigated based on a IP-based TONA handover architecture [5] which can be configured to support different levels or combinations of centralization. Here, we make a distinction between different radio resource management (RRM) distributions according to the levels of centralization. Accordingly, network-centralized RRM refers to RRM decisions made in a central access point controller (APC), network-distributed RRM refers to RRM decisions distributed between APs, and network-device distributed RRM refers to RRM decisions distributed between AP and stations.

### A. Predictive QoS Balancing Algorithm

In predictive QoS balancing (PQB) algorithm, the load metric is based on packet delay and packet loss rate which are derived by combining two analytical models, viz. Markov chain model to analyze the WLAN DCF operation and M/M/1/K queueing model to analyze the WLAN QoS performances. Here, we modify Zhai's model [2] to reflect the unbalanced load situation of an infrastructure-based VoWLAN in a wireline-to-wireless topology. The VoWLAN consists of one AP, $N-1$ WLAN stations and $N-1$ ethernet stations which are connected through a wireline backbone. When considering 2-way voice conversations between WLAN and ethernet stations, the traffic load flowing through the AP is $N-1$ times that of a WLAN station since AP transmits half of the voice traffic to WLAN stations. In addition, we introduce heterogeneous traffic between WLAN stations by considering voice codecs of different packetization intervals and packet length. We also model the freezing of backoff counter during times when medium is busy. The load metric is then used as upper bounds of admissible traffic load, which include the new and existing flows of an AP, in a centralized admission control to provision predictive QoS. We remark that these bounds are more proper as compared to those used in predictive load balancing algorithm since collision probability and queue characteristics of the AP are considered. However, PQB will generally result in higher complexity.

### B. Predictive Load Balancing Algorithm

In predictive load balancing (PLB) algorithm, the load metric is based on channel utilization which estimates the fraction of channel occupation time per observation interval. This load metric is widely used for both load and admission control algorithms due to its simplicity. Here, we implement PLB in a decentralized fashion as in [8]. Guaranteed QoS can be provisioned when both peak and mean channel utilization are used as upper bounds of admissible traffic load. Network utilization is usually acceptable when flows are smooth with constant bit rate (CBR) sources. However, when flows are bursty with variable bit rate (VBR) sources such guaranteed QoS inevitably results in low utilization. Higher network utilization can be achieved by relaxing these bounds to use mean channel utilization only. However, this means that only predictive QoS can be provisioned. Furthermore, the admission threshold for real-time (RT) flows is typically restricted to $80-90\%$. It is often argued that this buffer caters for variability of VBR sources and ensures that non real-time (NRT) flows can be accommodated within the buffered capacity. However, finding an optimal admission threshold in not trivial since the saturation point of WLAN depends on the proportion of traffic mixes e.g. RT vs. NRT flows and CBR vs. VBR sources. In other words, there will be a different impact on the network load *even* for the same average data rate. Hence, a better approach might be removing the admission threshold and relying on measurements of the existing flows to regulate input flows. Such measurements should be conservative by using historical knowledge of fluctuations in the network traffic.

TABLE I
COMPARISON OF DYNAMIC LOAD DISTRIBUTION ALGORITHMS.

| Attributes | Model-Based | Measurement-Based | |
|---|---|---|---|
| Algorithm Type | QoS balancing (PQB) | Load balancing (PLB) | QoS balancing (RQB) |
| Load Metric | Packet delay, packet loss | Channel utilization | Packet delay, channel utilization |
| Traffic Profiling | Mean arrival rates, collision probability, queue characteristics | Estimated peak and/or mean channel utilization | Measured packet delay, estimated mean channel utilization |
| Admission Control | Hard Limit | Hard Limit | Soft Limit |
| Centralization | Network-centralized RRM | Network-distributed RRM | Network-device distributed RRM |
| Information Exchange | Between APC-APs | Between APs | Between APC-AP-Stations |
| Stability Period | 5 Beacon Intervals | 5 Beacon Intervals | 20 Beacon Intervals |
| Utilization | Medium | Low | High |
| Complexity | High | Low | Moderate |

### C. Reactive QoS Balancing Algorithm

In reactive QoS balancing (RQB) algorithm, the load metric is based on measured packet delay and mean channel utilization which are utilized as upper bounds of admissible traffic load in network-device distributed RRM implementation found in [5]. These bounds are more relaxed as compared to the previous two algorithms, thus are referred as soft limits. Here, the mean channel utilization is used without imposing any admission threshold to RT flows. This essentially removes the hard limit and encourages higher network utilization. However, additional packet delay measurements need to be incorporated to account for the past network traffic variations. Accordingly, the measurements directly optimize the expected packet delay, making it adaptive to varying traffic conditions. This improves the flexibility of the admission control but at the expense of occasional violations, which limit it to provision predictive QoS, and moderate complexity. The network utilization gain would become more significant when there is a high degree of statistical multiplexing e.g. in broadband WLANs.

### D. Candidate Selection and Network Selection

To facilitate candidate selection, we quantify QoS requirements of stations as a function of two QoS metrics. Each QoS element is the ratio of the required QoS metric threshold and the measured QoS value. QoS satisfaction factor (QSF) is defined as the minimum between the two QoS elements,

$$QSF = \min_{i \in Links} \left[ \frac{PD^t}{PD_i^m}, \frac{PLR^t}{PLR_i^m} \right], \tag{1}$$

where $PD^t$ is packet delay threshold and $PLR^t$ is packet loss rate threshold while $PD_i^m$ is measured packet delay and $PLR_i^m$ is measured packet loss rate of $i$th links i.e. both uplink and downlink. $QSF < 1$ when QoS requirements of stations cannot be met. This condition is used by stations in all three dynamic load distribution algorithms to trigger QoS-based vertical handover.

The network selection in all three dynamic load distribution algorithms is based on the greedy approach. The reason being obtaining an optimal allocation of stations to available APs that maximize the composite network capacity is a combinatorial problem which is NP-hard. For PQB (PLB) algorithm, the AP which maximizes the difference between the estimated bounds and predefined QoS metric (load metric) thresholds is selected. For RQB algorithm, network selection is implemented according to [5] where AP with the highest network quality probability, which is based on the packet delay measurement, is selected. A Bayesian learning process is used to capture historical variations of network traffic conservatively, making it reliable for use in soft admission control.

TABLE II
TRAFFIC GENERATION PARAMETERS.

| Traffic Type | Packet Size (Bytes) | Inter-arrival (ms) | Avg. Data Rate (kbps) |
|---|---|---|---|
| G.711 | 80 | 10 | 64 |
| G.729 | 20 | 20 | 8 |
| G.723.1 | 24 | 30 | 6.4 |

## IV. COMPARATIVE PERFORMANCE EVALUATIONS

We simulate a hotspot of three 802.11b APs operating with data rate of $1Mbps$ under ideal channel conditions using OPNET™ Modeler® 14.5 wireless module. VoIP traffic are generated using heterogeneous voice codecs in Table. II and VBR source is simulated using ON-OFF model according to ITU [9]. We introduce an unbalanced load of five G.711, five G.729, five G.723.1 stations in BSS 1 and two G.711, two G.729, two G.723.1 stations in each of BSS 2 and BSS 3. The motivation is to examine the worst-case scenario when the total offered load approaches the composite network capacity of three BSSs. We assume no hidden terminals and exclude RTS-CTS mechanism. All stations are roaming to support handover events which are coordinated to one event at a time.

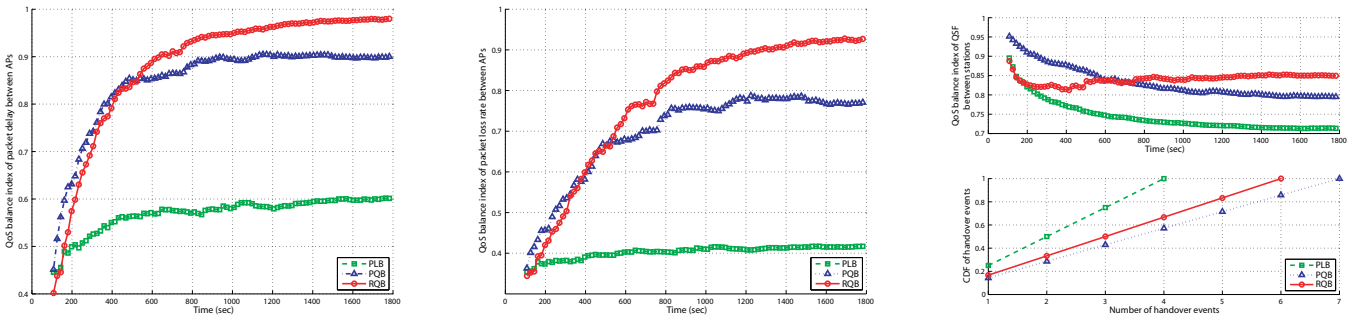
For performance evaluations, we adopt Jain's fairness index to quantify the effect of different dynamic load distribution algorithms on QoS fairness among APs (stations). Suppose $x_i$ is the QoS metric (QSF) of AP (station) $i$, then the QoS balance index (QBI) is defined as,

$$QBI(x) = \left(\sum\nolimits_i x_i\right)^2 \Big/ n\left(\sum\nolimits_i x_i^2\right), \tag{2}$$

where $n$ is the number of APs (stations). The QoS balance index $0 \leq QBI \leq 1$ is a continuous function which is independent of scale. It has a value of 1 when all AP (stations) have exactly the same QoS metric (QSF) and a value of $1/n$ when QoS metric (QSF) of APs (stations) are extremely unbalanced, which is 0 in the limit as $n \to \infty$.

### A. Results and Discussions

In this study, the key motivation is to quantify the state of balance between APs in terms of QoS metrics such as packet delay and packet loss rate, and between stations in terms of QSF defined in Eq. 1 when different dynamic load distribution




(a) QBI of packet delay. (b) QBI of packet loss rate. (c) QBI of QSF and CDF of handover events.

Fig. 1. QoS balance index between APs and stations, and CDF of handover events.

algorithms are deployed. According to the definition of Eq. 2, QBI should be close to one ideally to offer QoS fairness. We analyze our results starting from $100s$ ($0-100s$ is the warm-up period). From Fig. 1 (a) and (b), we observe that RQB outperforms both PQB and PLB by 4% (11%) and 54% (95%) in terms of QBI of packet delay (packet loss rate) between APs respectively. From 1 (c), we notice that both RQB and PQB have similar performance where the average QBIs of QSF between stations are 0.84 and 0.83 respectively. Both outperform PLB by 14%. However, PLB has the least number of handover events as compared to both RQB and PQB where PQB has an additional handover event as compared to RQB.

On the whole, both QoS balancing algorithms achieve better QoS fairness as compared to load balancing algorithm at the expense of an incremental number of handover events. QoS balancing algorithms exhibit better performance for two main reasons. First, the load metrics of both PQB and RQB contain at least one of the QoS metrics under study. This directly optimizes the expected packet delay and packet loss rate while the load metric of PLB is indirectly related to the investigated QoS metrics. Second, the load metric of PLB is based on mean channel utilization where the admission threshold is set to 90% of an AP maximum capacity. Since only BSS 1 is overloaded in the simulated scenario, the admission threshold creates an aggregate buffer of 20% preemptively in BSS 2 and BSS 3. This places a hard limit which prevents opportunistic exploitation of possibly spare capacity. Although this strategy attempts to protect existing flows, it inevitably results in higher blocking probability for incoming handover attempts. Hence, BSS 1 suffers sustained overloading which degrades the QoS fairness between APs and stations. We note that PQB also utilizes hard limit but admission threshold is not required. Hence, QoS fairness of PQB comes in-between RQB and PLB.

On the other hand, RQB also employs mean channel utilization in its load metric but relaxes the bounds by eliminating the admission threshold. Instead, it operates on a soft limit using packet delay measurement. Evidently, the salient advantage is achieving a higher network utilization by allowing exploitation of spare capacity opportunistically. Although there would be sporadic violations, this would be outweighed by the remarkable performance improvements such as in the case of RQB over PLB, where both are designed to provision predictive QoS. One interesting observation in Fig. 1 (c) is there exists a crossover point between PQB and RQB. This is because RQB relies on packet delay measurements and hence requires a longer stability period between handover events, to avoid ping-pong handovers, as compared to PQB. This crossover suggests that switching between load metrics may be beneficial in a truly heterogeneous multi-AP environment.

## V. Conclusion and Future Work

We evaluate the comparative performances between three dynamic load distribution algorithms, viz. predictive load balancing (PLB), predictive QoS balancing (PQB) and reactive QoS balancing (RQB) in terms of QoS fairness between APs and stations. The QoS metrics considered are packet delay and packet loss rate which are typically used to characterize the quality of VoIP traffic. Initial results show that RQB provides higher (significantly higher) network utilization and similar (much better) QoS fairness as compared to PQB (PLB). The results also suggest that all three algorithms depend largely on their load metrics. For future work, we plan to investigate the class of adaptive load distribution algorithm where load metrics can be dynamically adjusted according to prevailing system states and consider the impact of error-prone channels.

## References


[1] G. Bianchi and I. Tinnirello. Improving load balancing mechanisms in wireless packet networks. In *Proc. IEEE International Conference on Communications, 2002. ICC 2002*, volume 2, pages 891–895, 2002.

[2] H. Zhai, Y. Kwon, and Y. Fang. Performance analysis of IEEE 802.11 MAC protocols in wireless LANs. *Wirel. Commun. Mob. Comput.*, 4(8):917–931, 2004.

[3] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000.

[4] D. Malone, K.Duffy, and D. Leith. Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions. *IEEE/ACM Transactions on Networking*, 15(1):159–172, February 2007.

[5] E. H. Ong and J. Y. Khan. QoS provisioning for VoIP over wireless local area networks. In *Proc. 11th IEEE Singapore International Conference on Communication Systems, 2008. ICCS 2008*, pages 906–911, Guangzhou, China, November 2008.

[6] S. Garg and M. Kappes. Admission control for VoIP traffic in IEEE 802.11 networks. In *Proc. IEEE Global Telecommunications Conference, 2003. GLOBECOM '03*, volume 6, pages 3514–3518, December 2003.

[7] A. Balachandran, P. Bahl, and G. M. Voelker. Hot-spot congestion relief in public-area wireless networks. In *Proc. Fourth IEEE Workshop on Mobile Computing Systems and Applications, 2002*, pages 70–80, 2002.

[8] H. Velayos, V. Aleo, and G. Karlsson. Load balancing in overlapping wireless LAN cells. In *Proc. IEEE International Conference on Communications, 2004*, volume 7, pages 3833–3836, June 2004.

[9] ITU-T P.59. Artificial conversational speech. 1993.