

On Load Adaptation for Multirate Multi-AP Multimedia WLAN-Based Cognitive Networks

Eng Hwee Ong, Jamil Y. Khan

School of Electrical Engineering and Computer Science

University of Newcastle, Australia, NSW 2308

Email: enghwee.ong@studentmail.newcastle.edu.au, jamil.khan@newcastle.edu.au

Abstract—The widespread use of IEEE 802.11 WLAN and the potential to deliver QoS demanding multimedia contents, with the advent of 802.11n standard, will position itself as one of the introductory de-facto wireless access networks in the emerging cognitive networks. Existing WLANs can be augmented with cognitive functionality by the introduction of cognitive radios. However, the frequency-agile cognitive radio supports dynamic spectrum access to channels, which are heterogeneous and have largely different propagation characteristics, from diverse parts of the available frequency spectrum. Although WLAN supports link adaptation in practice, most of the existing WLANs are DCF-based which will give rise to the well-known rate anomaly problem where the long-term throughput of stations are penalized by the lowest data rate peer under multirate operation. Hence, deploying a WLAN-based cognitive network results in non-trivial radio resource management. In this paper, we introduce the novel concept of *load adaptation strategy* (LAS) to manage dynamic channel conditions associated with multirate multi-AP multimedia WLAN-based cognitive network in a single *unifying* QoS framework. Particularly, we show that our distributed LAS arbitrates optimal load distribution by maintaining a *QoS-balanced* system through QoS-based handovers in an opportunistic yet altruistic manner. Through simulations, we show that rate anomaly in multirate environment can be mitigated, statistical QoS guarantee can be provisioned for multimedia traffic with QoS fairness and system capacity can be maximized.

I. INTRODUCTION

In recent years, there is a plethora of independent radio access technologies (RATs), each supporting distinct coverage, mobility, data rates and QoS. In the future, additional RATs with other characteristics are expected to summate in a supplementary way. According to ITU's vision of *Optimally Connected, Anywhere, Anytime* published in Recommendation ITU-R M.1645, it aims at the integration of existing and new RATs to support data rates up to $100Mbps$ for high mobility applications and $1Gbps$ or more for nomadic mobility access. This stimulates the current trends towards convergence of heterogeneity where future wireless networks could be realized by a coalition of different RATs, each connected to a common IP-based core network to form a multi-RAT network.

To realize this vision, the context of cognition central to cognitive radio has been dilated to cognitive network. According to the recent IEEE 1900.1 standard [1], cognitive network is a composition of radio nodes subjected to cognitive functionality where the cognition process could take place either in the radio or in the higher layers or both. Cognitive network postulates the ability to optimize both user and network performances

as well as to bring about better utilization of radio resources by adopting cognitive functionality, particularly in the radio to exploit spectrum holes. Although cognitive radio can identify these spectrum holes, a parallel cognitive functionality is required in the higher layer to harness any effective use of these additional spectrum. Particularly, since the frequency spectrum is typically divided into multiple channels, the ramifications of dynamic spectrum access supported by cognitive radio manifests in the main challenge of managing these dynamically available heterogeneous channels from diverse parts of the spectrum with different propagation characteristics.

Within the research community, the IEEE 802.11 WLAN is envisioned as one of the introductory de-facto wireless access networks in the future cognitive network architecture, thanks to its pervasive deployments over many diverse environments. In fact, according to FCC [2], WLAN could be considered as cognitive radio since it operates with listen-before-talk access protocol together with dynamically changing transmission power and data rates to allow more efficient spectrum use. However, delivering QoS demanding multimedia applications over WLAN are not trivial, particularly in the context of a WLAN-based cognitive network. First, bulk of the existing WLANs are DCF-based which lack of service prioritization necessary to support multimedia traffic. Although the upcoming IEEE 802.11e standard supports service differentiation, it cannot guarantee strict QoS required by real-time services under heavy load without an appropriate admission control or load balancing mechanism [3].

Second, rate anomaly of DCF [4],[5] can occur when contending stations in the same BSS, with similar channel conditions, transmit packets of same size but with different data rates. Under such conditions, DCF preserves throughput fairness such that each station would yield approximately the same throughput, regardless of its own data rates. Rate anomaly of DCF is a direct consequence of throughput fairness which dictates equal probability of channel access. In other words, slower rate stations disadvantage higher rate stations by occupying more channel time to acquire approximately the same number of transmission opportunities. While DCF will affect channel capacity allocation and cause rate anomaly in the uplink, the AP queuing mechanism will affect channel bandwidth allocation and similarly induce rate anomaly in the downlink. Rate anomaly of AP queuing mechanism is attributed to varying channel conditions at different stations.

It will cause head-of-line blocking in DCF-based AP's FIFO queue when the channel quality towards the station of the head-of-line packet is degraded and AP performs link adaptation to transmit that packet at a lower rate.

Accordingly, rate anomaly will be unavoidable in multirate WLAN-based cognitive network since link adaptation are widely employed to combat varying channel conditions which could occur as a consequence of: (i) channel impairments in hotspot and indoor environments arising from frequent non-line-of-sight (NLOS) transmissions caused by structures and obstacles; and (ii) dynamic spectrum access supported by cognitive radio, resulting in heterogeneous channels operation. The impact of rate anomaly will escalate for WLAN-based cognitive network since cognitive radio can opportunistically access diverse channels which suffer very different amount of frequency dependent path loss, multipath effects and attenuations. Clearly, rate anomaly will create severe *capacity outages* when the offered traffic exceeds the downgraded system capacity, particularly, when the AP performs link adaptation to combat varying channel conditions at the stations. The negative impact of rate anomaly will be significant when two or more stations contend with diverse data rates *and* during heavy network load situations. Moreover, this impact would be magnified in the coexistence of 802.11b and 802.11g WLANs. Hence, a serious implication of rate anomaly is that it dilutes the benefits derived from any link adaptation techniques.

In this paper, we propose a novel LAS for multimedia service delivery in multirate WLAN-based cognitive network, where multi-AP are physically co-located, by mitigating rate anomaly. The basic idea is to exploit diversity within such multi-AP cognitive network to promote a *QoS-balanced* system by redistribution of load in an opportunistic yet altruistic manner. We note that this concept is applicable to any future multi-RAT environments. To the best of our knowledge, this is the first attempt to address rate anomaly in multirate WLAN-based cognitive network, where the serious implication of rate anomaly arising from link adaptation becomes imperative, from a single *unifying* QoS framework. The remainder of the paper is organized as follows. Section II discusses related work. Section III describes our proposed LAS. Section IV illustrates the simulation model. Section V presents the simulation results and Section VI concludes the paper.

II. RELATED WORK

Numerous solutions have been proposed to address rate anomaly prevalent in multirate environment. Most notable works address this problem from temporal fairness perspective [5],[6],[7] in order to overcome throughput fairness of DCF. Tinnirello *et al.* [6] have shown that EDCA channel access mechanism in the forthcoming IEEE 802.11e standard can provide temporal fairness by using transmission opportunity (TXOP) to maintain equal channel occupancy time for all contending stations. However, the authors showed that there is a tradeoff between achieving temporal fairness and overall system bandwidth efficiency as equalization of channel occupancy time requires fragmentation which introduces significant

overheads. Most importantly, the study also revealed that temporal fairness can only be guaranteed under uniform TXOP settings. However, this would be difficult to realize in practice since stations can select different fragmentation sizes without any fixed threshold, as opposed to DCF fragmentation rule. Moreover, the adoption of IEEE 802.11e standard by the industry remains questionable due to the high cost of replacing existing 802.11a/b/g hardwares for additional QoS support. Hence, the challenges in QoS provisioning for multimedia traffic delivery over DCF remains an open research issue.

Recently, Joshi *et al.* [7] presented a distributed time fair CSMA scheme to mitigate rate anomaly problem of DCF. The authors used the baseline property suggested in [5] to estimate a target throughput for each contending station. The key idea is that stations would adjust their minimum contention window to meet the estimated baseline throughput which has the intrinsic property of providing temporal fairness. However, the authors made several strong assumptions that limit pragmatic implementation. First, the authors assumed that every station has a priori knowledge of the number of contending stations which they conceded is non-trivial especially in multirate WLAN. Second, in their proposal, each station assumed that the transmission rate, packet size and packet errors experienced by all other stations are the same as itself. However, this scheme will fail to guarantee airtime fairness when traffic of different packet sizes exist in reality as baseline throughput is dependent on packet sizes [5].

Here, we take a completely different approach to mitigate rate anomaly from a QoS perspective. We first presented the QoS-inspired load optimization framework in [3] for a single rate WLAN. In this paper, we extend this framework and devise a harmonized solution to provision QoS for multimedia traffic in a dynamic multirate WLAN-based cognitive network, by introducing the notion of LAS. We show that the diversity of multi-AP cognitive network should be exploited, whenever possible, so that QoS requirements of multimedia stations can be better satisfied and not be penalized unnecessarily. Through LAS, stations with degrading QoS can be identified and transferred to a better quality AP in effort to maintain a QoS-balanced system. As soon as the station with degrading QoS is removed, link adaptation will increase the transmission rates of both the stations and AP which in turn recover the QoS and overall system capacity respectively.

Our contributions differ from related works in three significant ways: (i) we introduce the notion of *QoS balance*, which has intrinsic properties of mitigating rate anomaly, providing statistical QoS guarantee and maximizing overall system capacity, as criterion to quantify the state of balance in WLAN-based cognitive network; (ii) we exploit quadruple load metric consisting of packet delay, packet loss rate, signal-to-noise ratio and channel utilization for adaptation to dynamic network conditions prevalent in WLAN-based cognitive network; and (iii) our LAS, which is compatible with both DCF and EDCA access mechanisms, enables multimedia traffic delivery over WLAN-based cognitive network in a single unifying QoS framework.

III. PROPOSED LOAD ADAPTATION STRATEGY

The fundamental of our proposed LAS is based on *network-assisted discovery*, where source AP broadcasts measurement report of neighboring APs together with its own, over the terminal-oriented network-assisted (TONA) handover architecture as illustrated in Fig. 1. The estimated quadruple load metric (QLM) consisting of packet delay, packet loss rate, signal-to-noise ratio and channel utilization of an AP together with their channel number, geo-location and time of last handover event would be encapsulated in a packet as a measurement report and transmitted to the access point controller (APC) periodically. The APC would then collect these measurement reports from every AP in its subnet and facilitate exchanging of measurement reports between different subnets. The consolidated measurement reports of the source and neighboring APs would be disseminated from APC using a *cluster-based* broadcast. The cluster is defined as a group of ‘reachable’ APs bounded by the cluster radius r with respect to the geo-location of the source AP. The source AP only broadcasts measurement reports of that cluster so that stations need not monitor network conditions of distant APs.

The design philosophy of our LAS is based on the key principle of *QoS balance* by opportunistic yet altruistic exploitation of diversity within a multi-AP cognitive network, through a Bayesian learning network selection process [8]. The advantage of maintaining a QoS-balanced system is threefold. First, it mitigates rate anomaly in multirate environments through synergistic interactions of link adaptation and load adaptation. Second, it inherently provides statistical QoS guarantee for multimedia traffic. Third, it maximizes system capacity through better utilization of radio resources. Under the notion of QoS-balanced system, a handover will be triggered only if (i) QoS requirements of stations cannot be sustained; (ii) a better quality AP exists; and (iii) the requested handover will not disadvantage existing connections of the target AP. The context of disadvantage here refers to the situation when existing connections *fail* to meet their QoS requirements as a result of that handover. We achieve opportunistic yet altruistic exploitation when all the above conditions are met. We note that the first two conditions will preclude unnecessary handovers due to the reactive and opportunistic handover triggering approach.

As illustrated in Fig. 2, LAS is a cognitive functionality of the higher layer that resides in the MAC. It is responsible for synergistic interactions between PHY and MAC layers to exploit the benefits of both link adaptation and load adaptation on-demand. Specifically, load adaptation via vertical handover will be invoked when opportunistic yet altruistic exploitation is possible, otherwise link adaptation will be invoked. Furthermore, LAS provides synergistic interactions, between network-QoS and connection-QoS entities, within MAC layer to nominate candidate stations for handover based on the QLM which is shared between network entities through the TONA handover architecture. The network-QoS entity consists of service prioritization and admission control to deal with

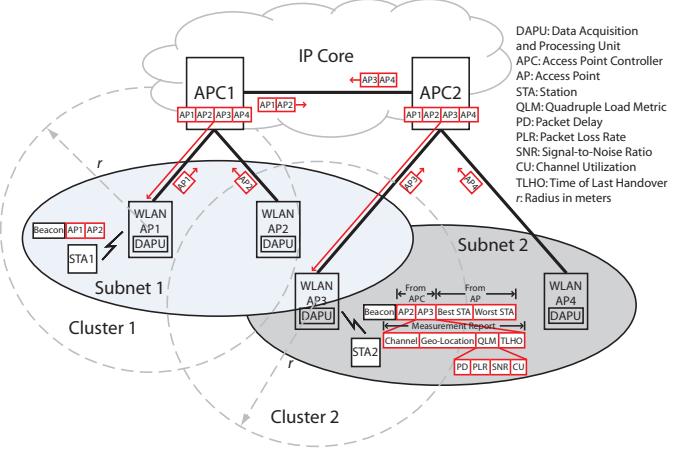


Fig. 1. Network-assisted discovery with cluster-based broadcast over TONA handover architecture.

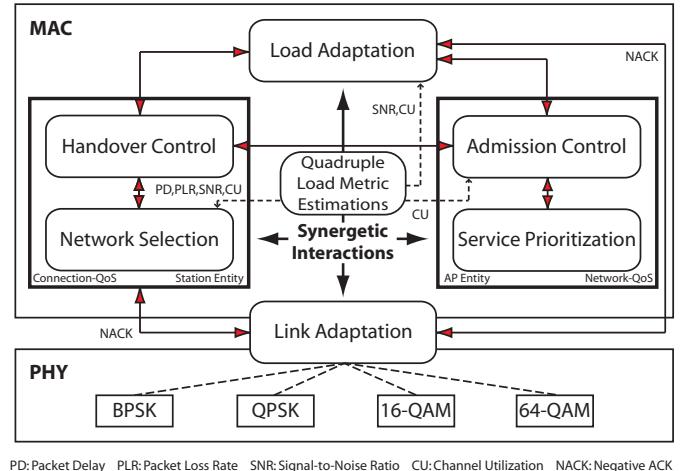


Fig. 2. Distributed LAS framework.

different user service profiles. On the other hand, connection-QoS entity consists of both network selection and handover control to deal with dynamic network conditions associated with network congestions, channel impairments and dynamic spectrum access.

The pseudo codes of the proposed LAS in both AP and station are given in Fig. 3 and 4 respectively. We employ negative acknowledgment (NACK) metric for detecting stations with degrading QoS and link adaptation procedures. The idea of NACK is similar to automatic rate fallback employed in Lucent WaveLAN-II where consecutive missed acknowledgment (MACK) signify bad channel conditions. However, MACK could also occur due to collisions hence we borrow the idea from [9] to use NACK instead for distinguishing between link errors and collisions. The load adaptation plays a central role by serving two main functions. First, it attempts to perform reactive load distribution by selecting the ‘best’ station for handover. Second, it identifies the ‘worst’ station with highest NACKs among stations with degrading QoS for handover, with priority over ‘best’ station, to mitigate rate anomaly by reducing the *capacity outage* time.

On Reception of Measurement Report from APC

```

Require: Periodic update of measurement report in each AP every beacon interval
1: AP extract QLM from measurement report
2: /* Compute total channel utilization in each AP */
3:  $CU_{total}^n \leftarrow \sum_{i \in STAs} CU_i^n, n \in APs$ 
4: /* Compute average channel utilization across APs */
5:  $CU_{ave} \leftarrow \frac{1}{n} \sum_n CU_{total}^n$ 
6: /* Determine available capacity for load distribution */
7:  $CU_{\Delta}^m \leftarrow CU_{ave} - CU_{total}^m, m \neq n$ 

On Link Adaptation
1: if NACK_RCVD then
2:    $NACK_i^n \leftarrow NACK_i^n + 1$ 
3: else // ACK_RCVD
4:    $ACK_i^n \leftarrow ACK_i^n + 1$ 
5: end if

On Load Adaptation
1: for all  $i \in STAs$  do
2:   if  $CU_{total}^n > CU_{ave}$  or  $SNR_n < SNR_{threshold}$  then
3:     /* Prevent load distribution to overloaded AP & overloading target AP */
Require:  $CU_{\Delta}^m - CU_i^n \geq 0, m \neq n$ 
4:     /* Nominate 'Best' STA for handover */
5:      $STA_{best}^n \leftarrow \min_{i \in STAs} (CU_{\Delta}^m - CU_i^n)$ 
6:   end if
7:   /* Nominate 'Worst' STA for handover */
8:    $STA_{worst}^n \leftarrow \max_{i \in STAs} (NACK_i^n)$ 
9: end for

On Admission Control
1: if  $CU_i^n + CU_{total}^n < CU_{max}$  then
2:   /* Invoke SERVICE_PRIORITIZATION for admitted connection */
3: end if

```

Fig. 3. Pseudo code of LAS algorithm in an AP.

We highlight that both ‘best’ and ‘worst’ stations correspond to the most aggressive traffic source associated with the AP. The ‘best’ station will be the traffic source with the highest channel utilization that can fit into the available capacity of the target AP. The ‘worst’ station will be the traffic source with the highest NACKs which turns out to be the traffic source with the highest channel utilization as well since the probability of NACKs increases with higher traffic arrival rate. This *highest channel utilization first out* (HCUFO) policy is attractive as it can essentially reduce the number of unnecessary handovers and result in long-term uniform distribution of aggressive traffic sources over the multi-AP cognitive network, which is beneficial from a load balancing perspective. As a final note, although handover is crucial for both mobility and load balancing, this work focuses on vertical handovers to redistribute load to a better quality or less loaded AP in an opportunistic yet altruistic manner.

IV. SIMULATION MODEL

The simulations for demonstrating the effectiveness of LAS are developed using OPNETTM Modeler[®] 14.0 with wireless module. We perform modifications to the existing DCF model to provide QoS support with handover control, network selection, admission control, service prioritization, link adaptation and load adaptation which are the focus of this study. An error-prone channel is considered by including shadow fading, multipath and variable path loss exponent to capture different propagation characteristic as a result of NLOS transmissions and/or dynamic spectrum access. Accordingly, we simulate shadow fading by varying the standard deviation of a zero-mean Gaussian distributed random variable

On Reception of Beacon from AP

```

Require: Periodic update of beacon information in each STA every beacon interval
1: STA extract QLM,  $T_{LHO}$  and HCUFO policy from beacon frame
On Initial Access
1: for all  $i \in$  initial access STAs do
2:   /* Invoke NETWORK_SELECTION */
3:   if  $QoS_n \geq QoS_{req,i}$  then
4:     /* Association with AP upon admission */
5:   end if
6: end for

On Synergetic Interactions between PHY and MAC
1: for all  $i \in$  handover STAs do
2:   /* Invoke LOAD_ADAPTATION */
3:   if  $QoS_m > QoS_n$  and  $PLR_n > 1\%$  and STA nominated,  $m \neq n$  then
4:     /* Invoke HANDOVER_CONTROL */
5:     while  $T_{current} - T_{LHO} < T_{stability}$  do
6:       /* Wait for stability period to elapse */
7:     end while
8:     if WORST_STA and BEST_STA are true then
9:       /* 'Worst' STA has priority */
10:    end if
11:    /* Reassociation with new AP upon admission */
12: else // Load adaptation conditions not met
13:   /* Invoke LINK_ADAPTATION */
14: end if
15: end for

```

Fig. 4. Pseudo code of LAS algorithm in a station.

uniformly $U[6, 9]$, multipath using the Exponential channel model, NLOS transmissions and effects of dynamic spectrum access by varying the path loss exponent uniformly $U[3, 4]$. We also model background noise by factoring in receiver noise figure simulated with a normal distribution $N(10, 0.1)$. We further model the network interface parameters according to Cisco Aironet 1130AG series AP and CB21AG series wireless client reference interface specifications to improve the realism of our simulations.

We simulate a hotspot with three 802.11g APs, operating with maximum data rate $54Mbps$. We assume that at least one legacy station is associated with each 802.11g AP since it is unlikely to operate in 802.11g only mode given the vast 802.11b deployments. However, the legacy station does not transmit any traffic so that all the system resources are available for 802.11g stations. We introduce a balanced load of three voice, three video, three FTP stations in each BSS. We then simulate channel variations in each BSS by introducing NLOS transmissions and effects of dynamic spectrum access. The channel variations are simulated according to binary representation of 0 to 7, i.e. 2^3 possible states, which result in either ‘high’ or ‘low’ SNR in each AP per 100s interval. Shadow fading and multipath are included for the entire simulation duration, and multimedia traffic sources are simulated according to Table I. We further assume no hidden terminals and exclude RTS-CTS mechanism from our simulation. However, CTS-to-self protection mechanism is used by 802.11g stations. We also incorporate MAC service data unit (MSDU) lifetime limit mechanism to discard MSDUs from the transmitter queue if they exceed the MSDU lifetime before successful transmission. The MSDU lifetime for voice, video and data packets are chosen as $50ms$, $150ms$ and $1s$ respectively. All stations are roaming capable to support handover events.

The QoS performance of our LAS is investigated from

TABLE I
TRAFFIC GENERATION PARAMETERS.

Traffic Type	Packet Size (Bytes)	Inter-arrival (ms)	Avg. Data Rate (kbps)
Voice-CBR (G.711)	80	10	64
Video-CBR	1000	125	64
Data-FTP (UL)	450	120	30
Data-FTP (DL)	1500	40	300

two critical aspects. First, 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,

$$QSF = \min_{i \in Class} \min_{j \in Links} \left[\frac{PD_i^t}{PD_{i,j}^m}, \frac{PLR_i^t}{PLR_{i,j}^m} \right], \quad (1)$$

where PD_i^t is packet delay threshold and PLR_i^t is packet loss rate threshold of i th service class while $PD_{i,j}^m$ is measured packet delay and $PLR_{i,j}^m$ is measured packet loss rate of j th links i.e. both uplink and downlink. $QSF < 1$ when QoS requirements of stations cannot be met.

Second, we adopt Jain's fairness index to quantify the effect of load adaptation on QoS fairness among stations. Suppose x_i is the QSF of station i , then the QoS balance index (QBI) is defined as,

$$QBI(x) = \left(\sum_i x_i \right)^2 / n \left(\sum_i x_i^2 \right), \quad (2)$$

where n is the number of 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 stations have exactly the same QSF and a value of $1/n$ when QSF of stations are extremely unbalanced, which is 0 in the limit as $n \rightarrow \infty$.

V. SIMULATION RESULTS

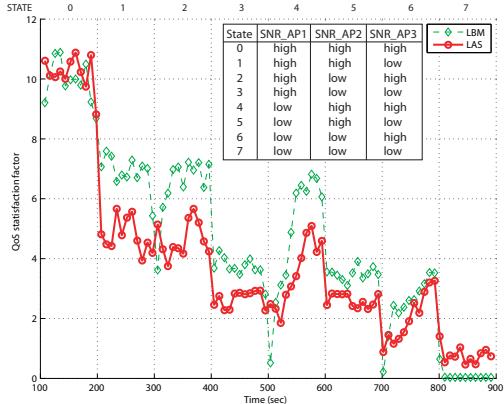
We first present the QoS performance of our LAS and aggregate throughput of system. We then compare it with a typical load balancing mechanism (LBM) implementation in [10] which utilizes channel utilization as the only load metric. According to the definitions in Eqs. 1–2, QSF should be greater than 1 and QBI should be close to 1, ideally to offer QoS guarantee and QoS fairness respectively. For QSF computation, the packet delay threshold for voice and video is 50ms, data is 300ms and packet loss rate threshold for all service classes is 2%. We analyze our results in eight states, which correspond to the simulated scenario, starting from 100s (0 – 100s is the warm-up period).

From Fig. 5(a), LAS has an average QSF of greater one except for state seven. LBM also has an average QSF of greater than one except for state seven and at the beginning of state four and six. In general, LBM has a higher average QSF than LAS except for state seven when all the APs have low SNR. This counterintuitive result can be explained from our design philosophy of avoiding unnecessary handovers when QoS requirements of stations can be supported. Since the traffic load distribution is the same with 68% load per AP,

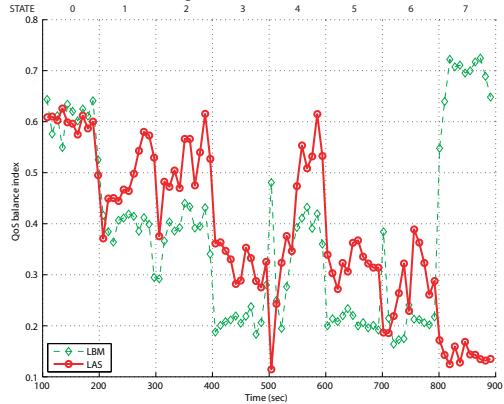
LBM does not trigger any handover as the load hysteresis δ is set to 10% and the channel utilization of all three APs are deemed balanced. However, we note that when the load hysteresis is zero, LBM has similar results in terms of average QSF and aggregate throughput but a 40% increase in handovers as compared to LAS. Although the load hysteresis could prevent unnecessary handovers, it is static and hence not responsive to channel variations. Evidently, the major pitfall of using channel utilization as a single load metric in LBM is the inability to manage channel variations which could arise due to NLOS transmissions and/or dynamic spectrum access in WLAN-based cognitive network. On the other hand, LAS enables handover of stations with degrading QoS to better quality AP in an opportunistic yet altruistic manner. Accordingly, the load of a good quality AP may be filled up to the maximum load threshold of 87%, thus the average QSF of LAS under such conditions will be lower than LBM. However, we emphasize that the QoS requirements of stations are not compromised since the average QSF of LAS is still greater than one, thanks to its altruistic design. On the contrary, LAS has higher average QSF than LBM during state seven even when all APs have low SNR since one of the APs has only 33% load while the remaining two APs have about 85% load each. It follows that the AP which is lightly loaded is not affected by low SNR, hence all the associated stations can still meet the QoS requirements which result in higher average QSF.

From Fig. 5(b), LAS achieves higher QBI than LBM for all states except for state seven, where LBM has higher QBI since QSF of all stations are equally bad. This observation is in direct contrast with that of Fig. 5(a) which suggests that tradeoffs exist between average QSF and QBI. For every increase in QBI achieved by LAS, there is a corresponding decrease in average QSF as compared to LBM. In other words, LAS trade QSF for QBI in order to maintain a QoS-balanced system. Although LBM has higher average QSF, it fails to provide any QoS fairness which means that there is a huge disparity in the QoS or throughput between stations of the same service class. We have excluded the results on QBI of stations per service class due to space but we note that similar trends to Fig. 5(b) are observed.

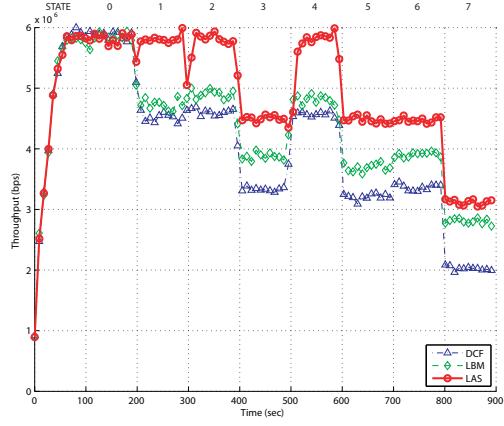
Finally from Fig. 5(c), LAS attains the highest aggregate throughput as compared to LBM and DCF across all states. We note that DCF serve as a baseline for our comparative study since both LAS and LBM are essentially DCF-based. Apparently, both LBM and DCF suffer marked throughput degradation as a consequence of rate anomaly in the downlink which causes head-of-line blocking at the AP's FIFO queue. This subsequently extends to buffer overflow and excessive real-time packets being dropped due to MSDU lifetime expiry. In this simulation, rate anomaly in the downlink has a more dominant effect than uplink since infrastructure-based WLAN is downlink limited and DCF has essentially a single FIFO queue in the MAC layer. We can also conclude that although LBM (with link adaptation) yields better throughput than DCF (without link adaptation) under varying channel conditions, the throughput enhancement is *rate-limited* by link adaptation.



(a) Average QoS satisfaction factor.



(b) QoS balance index.



(c) Aggregate throughput of system.

Fig. 5. Simulation results for average QSF and QBI of stations, and aggregate throughput of system.

Clearly, by maintaining a QoS-balanced system, LAS provides better utilization of radio resources by redistribution of load in an opportunistic yet altruistic manner. This has intrinsic properties of mitigating rate anomaly in multirate environment, providing statistical QoS guarantee and maximizing overall system capacity. Another interesting observation is that LAS has a throughput gain of 20% over LBM in states zero, one, two and four when more than 2/3 of the APs have high SNR. On the other hand, LAS has throughput gain of 12.5% over LBM in states three, five and six when only 1/3 of the APs have high SNR. Clearly, this suggests that the throughput and

QoS performance gain of LAS will scale with the diversity of increasing number of APs since the probability of all APs with low SNR at the same time is remote. We argue that in future heterogeneous wireless networks, multi-AP deployments in excess of three APs will become a reality and diversity should be exploited to harness higher overall system capacity.

VI. CONCLUSION

We present a novel LAS, which is compatible with both DCF and EDCA channel access mechanisms, in a single *unifying* QoS framework for multirate multi-AP multimedia WLAN-based cognitive network. This work highlights that *QoS balance* has salient traits of mitigating rate anomaly in multirate environments, providing statistical QoS guarantee for multimedia traffic and maximizing system capacity through the better use of radio resources. Hence, we advocate QoS balance as criterion to quantify the state of balance in WLAN-based cognitive network. We find that typical load balancing with a single load metric such as channel utilization is unable to cope with dynamic channel conditions. By employing quadruple load metric, we show that LAS enables synergistic interactions of link adaptation and load adaptation to overcome these dynamic channel conditions in a self-adjusting manner. Finally, we argue that diversity of multi-AP cognitive network should be exploited, whenever possible, to promote better utilization of radio resource. How to exploit all possible diversities is worth further study.

REFERENCES

- [1] IEEE 1900.1-2008. IEEE standard definitions and concepts for dynamic spectrum access: Terminology relating to emerging wireless networks, system functionality, and spectrum management. September 2008.
- [2] FCC 05-57. Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies. In *ET Docket No. 03-108*. Federal Communications Commission, March 2005.
- [3] E. H. Ong and J. Y. Khan. A unified QoS-inspired load optimization framework for multiple access points based wireless LANs. In *Proc. IEEE Wireless Communications and Networking Conference, 2009. WCNC 2009*, pages 1–6, Budapest, Hungary, April 2009.
- [4] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda. Performance anomaly of 802.11b. In *Proc. IEEE INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 836–843, March/April 2003.
- [5] G. Tan and J. Guttag. Time-based fairness improves performance in multi-rate wlans. In *Proc. ATC '04: Proceedings of the annual conference on USENIX Annual Technical Conference*, pages 23–23, Berkeley, CA, USA, 2004. USENIX Association.
- [6] I. Tinnirello and S. Choi. Temporal fairness provisioning in multi-rate contention-based 802.11e WLANs. In *Proc. World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a*, pages 220–230, June 2005.
- [7] T. Joshi, A. Mukherjee, Younghwan Yoo, and D. P. Agrawal. Airtime fairness for IEEE 802.11 multirate networks. *IEEE Transactions on Mobile Computing*, 7(4):513–527, April 2008.
- [8] E. H. Ong and J. Y. Khan. Dynamic access network selection with QoS parameters estimation: A step closer to ABC. In *Proc. IEEE Vehicular Technology Conference, 2008. VTC Spring 2008*, pages 2671–2676, Marina Bay, Singapore, May 2008.
- [9] Q. Pang, V. C. M. Leung, and S. C. Liew. A rate adaptation algorithm for IEEE 802.11 WLANs based on MAC-layer loss differentiation. In *Proc. Broadband Networks, 2005. BroadNets 2005. 2nd International Conference on*, pages 659–667, October 2005.
- [10] 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.