Distributed Radio Resource Usage Optimization of WLANs based on IEEE 1900.4 Architecture

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—IEEE 1900.4 is the recently approved standard which specify a policy-based radio resource management (RRM) framework where the decision-making process is distributed between the network-terminal entities. By exploiting the cooperative exchange of context information between the network-terminal entities, the standard aims to facilitate the optimization of radio resource usage to improve the overall composite capacity and QoS of wireless networks in a multiple radio access technologies (multi-RAT) landscape. In this paper, we demonstrate how IEEE 1900.4 standard can be employed to legacy radio access network, in particular, a multiple access point (multi-AP) based WLAN to effectuate distributed radio resources usage optimization, corresponding to one of the three use cases defined in the standard through a load adaptation policy (LAP). Through simulations, we show that overall composite capacity and QoS in terms of QoS balance can be improved by 15% and 23% respectively and the total number of handover can be reduced by 72% with IEEE 1900.4 RRM as compared to a network-distributed RRM architecture.

I. INTRODUCTION

Future wireless networks would be radically different from todays independent RATs through the widely accepted notion of convergence in heterogeneity. 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 integration of different RATs over a common IP-based core network to form a multi-RAT landscape. In recent times, there are vested interests in cognitive network that applies cognitive principles to other layers of the protocol stack. The motivation is to exploit both cognitive functionality and cooperation [1] to improve user and network performances through better utilization of radio resources in the presence of heterogeneity, and convergence of networks, terminals and services in future wireless networks.

Recently approved IEEE 1900.4 standard [2], which defines a distributed decision-making process between the network-terminal entities, will realize this ultimate goal of improving overall composite capacity and QoS of future wireless networks in a multi-RAT environment. Prior to the standardization of IEEE 1900.4, we have developed a QoS-inspired load optimization framework [3], where network load is redistributed across a single rate multi-AP WLAN in self-adjusting manner to maintain a QoS-balanced system. The key idea of broadcasting network context information to terminals, which subsequently select an AP that can ‘best’ fulfill their QoS requirements dynamically, through a terminal-oriented network-assisted (TONA) handover architecture [4],[1] is similar to the IEEE 1900.4 system architecture.

In this paper, we relate our TONA handover architecture to the IEEE 1900.4 standard and propose a LAP to implement the distributed radio resource usage optimization use case in Annex A, §A.3 of [2] for a multirate multi-AP WLAN. To the best of our knowledge, this is one of the first studies to provide insights on benefits of the recent IEEE 1900.4 RRM architecture [2] based on the distributed radio resource usage optimization use case as compared to the network-distributed RRM architecture described in [5] and implemented in [6]. The remainder of the paper is organized as follows. Section II gives an overview of the IEEE 1900.4 standard. Section III relates our TONA handover architecture to the IEEE 1900.4 standard and then illustrates our proposed LAP. Section IV introduces the simulation model. Section V discusses the simulation results and Section VI concludes the paper.

II. IEEE 1900.4 STANDARD OVERVIEW

The IEEE 1900.4 standard specifies an architecture which supports distributed decision-making between the network-terminal entities to improve the overall composite capacity and QoS of a composite wireless network (CWN). According to the standard, CWN refers to a coalition of radio access networks (RANs), each connected through a packet (IP) based core network with the deployment of IEEE 1900.4 entities. To realize this, the IEEE 1900.4 defines three use cases: (i) distributed radio resource usage optimization for legacy RAN where frequency bands are fixed and RAN are not reconfigurable; (ii) dynamic spectrum assignment where frequency bands are dynamically assigned to RANs to optimize spectrum usage; and (iii) dynamic spectrum sharing where frequency bands are fixed but can be shared and/or used dynamically by RANs and terminals. Here, we focus on the first use case and show how our TONA handover architecture and proposed LAP can be easily extended to the other two use cases.

To implement the defined use cases, the IEEE 1900.4 defines the system architecture as illustrated in Fig. 1 comprising of seven entities. Four entities are defined on the network side. The operator spectrum manager (OSM) is used to reflect the operators’ control over the spectrum assignment policies.
for the network reconfiguration manager (NRM). The NRM manages the CWN and terminals by generating radio resource selection policies to guide the terminal through the optimized radio resource allocations. The RAN measurement collector (RMC) acquires RAN context information for the NRM. The RAN reconfiguration controller (RRC), which acts upon RRM requests, is responsible for the reconfigurations of RANs. We note that the NRM, RMC and RRC may be implemented in a distributed manner.

Three entities are defined on the terminal side. The terminal reconfiguration manager (TRM) manages the terminal by making the final decision regarding radio resource allocation within the bounds of guiding policies defined by the NRM, user preferences and the available context information. The terminal measurement collector (TMC) acquires terminal context information for the TRM. The terminal reconfiguration controller (TRC), which acts upon TRM requests, is responsible for the reconfigurations of terminals. Of the seven defined entities, both NRM and TRM are the key decision-making entities where exchanges of RAN context information and terminal context information, including the dissemination of network policies are carried out by the radio enabler (RE). In this paper, we focus on the implementation aspects of the LAP associated with the NRM and TRM. Readers are referred to [7] for development aspects of the RE.

III. TONA HANDOVER ARCHITECTURE AND PROPOSED LOAD ADAPTATION POLICY

The main idea of the IEEE 1900.4 is enabling terminals to participate in the decision-making process autonomously while adhering to some policies and constraints imposed by the network. Our proposed LAP is based on such distributed decision-making process between network-terminal entities by leveraging on the TONA handover architecture to exploit cooperative exchange of context information as illustrated in Fig. 2. Accordingly, the TONA handover architecture supports: (i) network-assisted discovery where source AP broadcasts network context information of neighboring APs together with its own and the recommended LAP; and (ii) terminal-oriented decision where terminals make final RRM decision in selecting an AP that can ‘best’ fulfill the user preferences and QoS requirements while operating within bounds of the recommended LAP. In effect, our TONA handover architecture corresponds to the IEEE 1900.4 system and functional architectures in the following aspects.

First, the measurement report, which consists of the estimated quadruple load metric (QLM), time of last handover event, channel number and geo-location of the AP, is analogous to the RAN context information. Second, QLM comprises of packet delay, packet loss rate, signal-to-noise ratio and channel utilization. These are estimated with the data acquisition and processing unit (DAPU) [4],[1] which is analogous to both RMC and TMC. Third, the measurement report is transmitted to the access point controller (APC), which is analogous to the NRM, periodically. Fourth, the APC would collect these measurement reports from every AP in its subnet and facilitate exchanging of measurement reports between different subnets through the IP based core network should multiple NRMs exist as stated in the standard. Fifth, the consolidated measurement reports of the source and neighboring APs would be disseminated from APC using a cluster-based broadcast based on geo-location of both the AP and terminals (or TRMs) as suggested in the standard. We defined the cluster as a group of ‘reachable’ APs and terminals bounded by the cluster radius r with respect to geo-location of the source AP. The motivation is to dispense the terminal from monitoring network conditions of distant APs which are ‘unreachable’. This is in addition to the standard which only requires AP to broadcast measurement reports to selected terminals within the geo-localized cluster. Last, the radio enabler would be mapped to the beacon (in-band channel) of the source AP for broadcast as one of the two options specified by the standard.

The design philosophy of our LAP is based on the key principle of QoS balance by opportunistic yet altruistic exploitation of diversity within a multi-AP WLAN. Under the notion of QoS-balanced system, a handover will be triggered only if (i) QoS requirements of stations\(^1\) cannot be sustained;
First, channel impairments are apparent in legacy WLAN deployments at hotspot and indoor environments where struct-

(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.

The algorithm of our proposed LAP in the NRM and load adaptation decision in the TRM are illustrated in Fig. 3 and 4 respectively. The LAP is responsible for synergetic interactions between PHY and MAC layers and within MAC layer to exploit the benefits of both link adaptation and load adaptation on-demand. Through the TONA handover architecture, QLM of each AP is exchanged between network-terminal entities to facilitate joint optimization of radio resource usage. The network-QoS entity, which consists of service prioritization and admission control to deal with different user service profiles, resides in each AP (or RAN). On the other hand, connection-QoS entity, which consists of network selection and handover control to deal with dynamic network conditions associated with network congestions, channel impairments and dynamic spectrum assignment/sharing, resides in each terminal. Bayesian learning is employed in the NRM for both PD and PLR to obtain the required posterior distributions for network selection [4],[1] in the TRM.

The LAP plays a central role by serving two main functions. First, it attempts to perform load balancing by selecting the ‘best’ station for planned handover. Second, it identifies the ‘worst’ station with highest NACKs among stations with degrading QoS for unplanned handover, with priority over ‘best’ station, to mitigate rate anomaly [3] prevalent in multirate WLAN environment. 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 also corresponds to the traffic source with the highest channel utilization 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 WLAN, which is beneficial from a load balancing perspective.

To this end, we have given a detailed exposition on the implementation of the distributed radio resource usage optimization use case based on our proposed LAP. We note that inter-operator cooperation is not the focus in this paper and hence assume that such cooperation exists. Without loss of generality, our TONA handover architecture and proposed LAP can be easily extended to the other use cases of dynamic spectrum assignment and dynamic spectrum sharing. To realize this, we require two additional IEEE 1900.4 entities, viz. OSM and RRC as well as two additional functional blocks (in dashed lines) in the NRM, viz. spectrum assignment evaluation, and network reconfiguration decision and control, all on the network side. Nevertheless, the effects of both dynamic spectrum assignment and dynamic spectrum sharing are simulated as an initial study.

Fig. 3. LAP implementation in the NRM.

Fig. 4. Load adaptation decision implementation in the TRM.

IV. SIMULATION MODEL

The rationale of this simulation study is to capture the inherent practical issues associated with legacy WLAN deployment and dynamic spectrum assignment/sharing capabilities supported by reconfigurable RANs/terminals, which correspond to the three defined use cases in IEEE 1900.4 standard. First, channel impairments are apparent in legacy WLAN deployments at hotspot and indoor environments where struc-
tures and obstacles cause frequent non-line-of-sight (NLOS) transmissions. Second, dynamic spectrum assignment/sharing results in opportunistic access to diverse channels of the available frequency spectrum which will have largely different propagation characteristics. Third, although dynamic spectrum assignment/sharing exploits spectrum holes of primary users opportunistically, the secondary users are mandated to utilize the spectrum only when its transmission does not interfere with that of the primary users. This hard requirement means that the availability of spectrum and the corresponding system capacity would essentially be time-varying, depending on the load of the primary users. These issues imply that a suitable LAP must be in place to manage fruitful utilization of heterogeneous channels, prevent any serious QoS degradation when capacity ceases to exist and to maximize system capacity when extra capacity is available.

The OPNET\textsuperscript{TM} Modeler\textsuperscript{®} 14.0 with wireless module is used for our simulation study. We modified 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 assignment/sharing. Accordingly, we simulate shadow fading by varying the standard deviation of a zero-mean Gaussian distributed random variable uniformly $U[6, 9]$, multipath using the Exponential channel model, NLOS transmissions and effects of dynamic spectrum assignment/sharing 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 $54\text{Mbps}$. 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 AP. We then simulate (i) capacity variations by introducing the effects of dynamic spectrum assignment/sharing in each AP; and (ii) channel variations by introducing NLOS transmissions and affects of dynamic spectrum assignment/sharing in each BSS. The capacity and channel variations are simulated according to Table I. For states 1 – 7, we assume that AP with reduced capacity of $6\text{Mbps}$ has high SNR of $35\text{dB}$ due to its more robust modulation and coding scheme while AP with higher capacity of $54\text{Mbps}$ may have either high SNR of $35\text{dB}$ or low SNR of $15\text{dB}$ depending on propagation characteristic. The motivation is to investigate the consequence of load balancing under varying capacity and channel conditions. Shadow fading and multipath are included for the entire simulation duration, and multimedia traffic sources are simulated according to Table II. We further assume no hidden terminals and exclude RTS-CTS mechanism. 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 $50\text{ms}$, $150\text{ms}$ and $1\text{s}$ respectively. All stations are roaming capable to support handover events without multi-homing capability.

The QoS performance of our LAP is investigated from 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,

$$QSF = \min_{i \in \text{Class}} \min_{j \in \text{Links}} \left[ \frac{PD_{i}^{t}}{PD_{i,j}^{m}} \frac{PLR_{i}^{t}}{PLR_{i,j}^{m}} \right] ,$$

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 QoS fairness among stations. Suppose $x_{i}$ is the QSF of station $i$, then the QoS balance index (QBI) is defined as,

$$QBI(x) = \left( \sum_{i} x_{i} \right)^{2} / n \left( \sum_{i} x_{i}^{2} \right),$$

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 $n \to \infty$.

V. SIMULATION RESULTS

We evaluate the effectiveness of the IEEE 1900.4 RRM, based on the proposed LAP, in terms of aggregate throughput.

| Table I | CAPACITY AND CHANNEL VARIATIONS. |
|---|---|---|---|---|
| State | AP1 | AP2 | AP3 | AP1 | AP2 | AP3 |
| 1 | 54 | 54 | 54 | 35 | 35 | 15 |
| 2 | 54 | 54 | 6 | 35 | 15 | 35 |
| 3 | 54 | 6 | 54 | 15 | 35 | 15 |
| 4 | 54 | 6 | 6 | 15 | 35 | 35 |
| 5 | 6 | 54 | 54 | 35 | 15 | 15 |
| 6 | 6 | 54 | 6 | 35 | 35 | 15 |
| 7 | 6 | 6 | 54 | 35 | 35 | 35 |
| 8 | 6 | 6 | 6 | 15 | 35 | 35 |

| Table II | 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 |
of system, QBI of stations’ QSF, average QSF of stations and total number of handover events as shown in Figs. 5(a)–(d) respectively. We then compare it with the network-distributed RRM, based on a typical load balancing mechanism (LBM) implementation in [6] which utilizes absolute network load, particularly, 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 50 ms, data is 300 ms 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).

In general, we observe that the aggregate throughput and QBI of LAP outperform LBM by 15% and 23% respectively as shown in Fig. 5(a) and 5(b) for the simulated scenario. However, LBM has a higher average QSF than LAP in state one, two, five and six as shown in Fig. 5(c). This counterintuitive result is a direct consequence of our design philosophy to avoid unnecessary handover when QoS requirements of stations can be supported. In state one, LBM does not trigger any handover as the channel utilization of all three APs is balanced. However, LAP is aware of the low SNR in AP3 and triggers five handovers to redistribute load from AP3 to both AP1 and AP2 which have better SNR. Although LBM has a higher average QSF than LAP, the aggregate throughput of LAP outperforms LBM by 22% and the QBI of LAP outperforms LBM by 36%.

In state two, LBM triggers five handovers from AP3 to both AP1 and AP2 since the capacity of AP3 is reduced. However, we note that load balancing based on channel utilization as a single load metric is catastrophic since the load in this case is ‘wrongly’ transferred to AP2 with low SNR. This has a strong negative impact on both aggregated throughput and QoS. On the other hand, LAP triggers four handovers to transfer load from AP2 to AP3 instead since it is aware that AP3 has higher SNR than AP2 and can still accept some additional load after some of its original load has been redistributed in state one. Although LBM has a higher average QSF than LAP, the aggregate throughput of LAP outperforms LBM by 21% and the QBI of LAP outperforms LBM by 17%.

In states three and five, both LAP and LBM have an average QSF of less than one. The compromise of QSF is unavoidable as these two states simulate the extreme case where capacity of one AP is reduced and the alternative APs have low SNR. LBM, agnostic of the channel conditions, continues to transfer load to APs with low SNR by triggering eleven handovers in state three and twelve handovers in state five. On the contrary, LAP does not transfer any load since it is aware of the low SNR in the alternative APs which inhibits any exploitation of diversity. Under such conditions, we highlight that the aggregate throughput of LAP still outperforms LBM by 12% and the QBI of LAP also outperforms LBM by 52% in state three while similar performances are observed in state five.

States four and six correspond to the scenario where the capacity of two APs is reduced and the alternative AP suffers low SNR. LBM triggers eight handovers from AP2 and AP3 to AP1 with low SNR in state four and triggers eleven handovers from AP1 and AP3 to AP2 with low SNR in state six. On the other hand, LAP does not trigger any handover in state four while one handover is triggered from AP2 to AP1 in state six. Although LBM has a higher average QSF than LAP in state six, the aggregate throughput of LAP outperforms LBM by 33% and the QBI of LAP outperforms LBM by 26%. As for state four, the aggregate throughput of LAP outperforms LBM by 30% and the QBI of LAP outperforms LBM by a remarkable 66%.

State seven corresponds to the scenario where the capacity of two APs is reduced and all APs have high SNR. LBM triggers five handovers from AP1 and AP2 to AP3 but this time AP3 has high SNR. On the other hand, LAP does not trigger any handover since QoS requirements of stations associated with AP1 and AP2 can still be met. This is the very reason why the QBI of LBM is higher than LAP since it tries to perform load balancing proactively while LAP takes the reactive approach to trigger handover only when QoS requirements of stations cannot be met. We note that, in this case, the aggregate throughput and the average QSF of both LAP and LBM are similar. More importantly, we emphasize that the QoS requirements of stations are not compromised since the average QSF of LAP is still greater than one. Finally, state eight represents the scenario where the capacity of all APs is reduced and one of the APs has low SNR. Under such conditions, the average QSF of LAP outperforms LBM by 26%, the aggregate throughput of LAP outperforms LBM by 5% and the QBI of LAP outperforms LBM by 29%.

Although LBM is balanced from the network perspective in terms of channel utilization as shown in Fig. 5(b), it fails to provide QoS fairness in six of the eight simulated states when the target AP has low SNR. We have excluded the results on QBI of stations’ QSF per service class due to space but we note that similar trends to Fig. 5 (b) are observed. In essence, low QBI of LBM means that there is a huge disparity in the QoS or throughput between stations of the same service class. This has serious implications on the three defined use cases in the IEEE 1900.4 standard which focuses on load balancing of available spectrum. Our results underpin that any feasible spectrum load balancing shall be QoS-aware which can be realized by the exchange of QoS context information such as the QLM. In addition, we advocate that QoS balance should be used as criterion to quantify the state of balance in CWN from the terminal perspective, with the notion of class-based QoS i.e. QSF of stations per service class, instead of from the network perspective based on absolute network load. Moreover, QoS balance has intrinsic property of providing statistical QoS guarantee for all states except three and five. Evidently, the QoS-agnostic LBM is incapacitated by channel variations which could arise due to NLOS transmissions in legacy WLAN and/or dynamic spectrum assignment/sharing supported by reconfigurable RANs/terminals. As a final note, we highlight that there are a total of fifty-four handovers
triggered by LBM but only fifteen handovers triggered by LAP as shown in Fig. 5(d). This translates to a massive 72% reduction in handover while achieving high overall composite capacity and maintaining a QoS-balanced system.

VI. CONCLUSION

We propose a LAP by leveraging on our TONA handover architecture which is similar to the recent IEEE 1900.4 standard in many aspects. The proposed LAP implements the distributed radio resource usage optimization use case, and is easily extensible to the dynamic spectrum assignment and dynamic spectrum sharing use cases. We show that the IEEE 1900.4 RRM can effectively exploit the cooperative exchanges of context information, such as the QLM, between network-terminal entities to facilitate the coordinated use of radio resources and QoS awareness, thus harnessing overall composite capacity and QoS improvements. We also find that the network-distributed RRM which performs load balancing based on a single load metric such as channel utilization is QoS-agnostic and will be incapacitated by channel variations which could arise due to NLOS transmissions in legacy WLAN and/or dynamic spectrum assignment/sharing supported by reconfigurable RANs/terminals. As a consequence, the IEEE 1900.4 RRM outperforms network-distributed RRM in terms of aggregate throughput of system by 15%, QoS balance by 23% and handover reduction by 72%. Finally, we advocate that QoS balance should be employed as criterion to quantify the state of balance in IEEE 1900.4 based CWN from the terminal perspective, with the notion of class-based QoS, instead of from the network perspective based on absolute network load. Moreover, QoS balance has the intrinsic property of providing statistical QoS guarantee for multimedia traffic while maximizing the overall composite capacity.

REFERENCES