Whittier, A., Kulkarni, P., Cao, F., & Armour, S. (2016). Mobile data offloading addressing the service quality vs. resource utilisation dilemma. In IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2016. [16555981] (Proceedings of PIMRC). Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/PIMRC.2016.7794869

Peer reviewed version

Link to published version (if available):
10.1109/PIMRC.2016.7794869

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via IEEE at http://ieeexplore.ieee.org/document/7794869/. Please refer to any applicable terms of use of the publisher.

Mobile Data Offloading addressing the Service Quality vs. Resource Utilisation Dilemma

Alicia Whittier*†, Parag Kulkarni*, Fengming Cao*† and Simon Armour†
*Telecommunications Research Laboratory, Toshiba Research Europe Ltd., Bristol, UK
†Dept. Of Electrical & Electronic Engineering, University of Bristol, UK
Email: {alicia.whittier, parag.kulkarni, fengming.cao}@toshiba-trel.com, simon.armour@bristol.ac.uk

***Abstract*—Mobile operators are integrating WiFi into their cellular networks as a way to address the issue of mobile data deluge. Although prominently used for its high data rates and unlicensed frequency band of operation, WiFi's contention based medium access does not guarantee that a user's quality of service will necessarily be met, especially if mobile devices simply 'offload' to WiFi in preference to cellular. A network selection algorithm offers a way to enhance user experience, by always selecting the best radio access network for the user. In the literature, several techniques have been applied to the network selection problem. While most proposals aim to optimise quality of service for the user, the impact on resource utilisation for the networks is often neglected. This work proposes a network selection algorithm which attempts to find a good trade-off between user quality of service and efficient resource utilisation and elaborates on its performance evaluation through a simulation based study. Findings indicate that the proposed algorithm is able to provide enhanced QoS to the user as compared to conventional algorithms in terms of achieving higher throughput, lower delay as well as a higher overall system performance.**

***Keywords*-3GPP LTE; WiFi 802.11; mobile data offloading; QoS**

I. Introduction & Related Work

Operators have been evolving their networks to provide increased data speeds and deploying small cell solutions to meet with the increased demand. They have also been offloading traffic to WiFi access points (APs) since Wi-Fi offers high data rates and has the advantage of operating in an unlicensed frequency band. However, when offloading users to WiFi, maintaining the quality of service (QoS) for the user remains a big challenge [1]. This ultimately depends on the network selection algorithm which aims for choosing the best network for the user given the choice of radio access technologies (RATs) in its coverage area. At the same time, radio resources should also be used efficiently.

Provisions for cellular/WLAN interworking have existed for a while, providing options for both tightly and loosely coupled architectures to enable mobile operators to take advantage of WLAN offloading, either through deploying their own WiFi APs or through partnering with a WLAN Internet Service Provider (ISP). To assist user equipment (UEs) in discovering Wi-Fi APs and understanding their capabilities before associating with it, the Wi-Fi Alliance introduced its Hotspot 2.0 specification, which also incorporates the IEEE 802.11u amendment (interworking with external networks). IEEE 802.11u allows an AP to advertise extra information about its WLAN network to the UE (via its beacon frames) but also allows a UE to query an AP, should other information be required. The main purpose of the Hotspot 2.0 specification is to assist with seamless roaming between the cellular and WLAN network by making the process of discovering, selecting, authenticating and connecting to the WLAN network unnoticeable to the user [2].

To provide operators with some control over which WiFi APs could be discovered and selected by the UE, 3GPP has also introduced its Access Network Discovery and Selection Function (ANDSF) [3]. It includes a WLAN selection policy which allows an ANDSF server located in the mobile core network to provide a policy to the UE, with rules for discovering and selecting Wi-Fi APs. ANDSF and Hotspot 2.0 actually complement each other and could be envisaged to operate in unison whereby an operator's ANDSF policy rule could be applied to a UE, allowing it to only associate with the WiFi AP provided the Hotspot 2.0 measurements/information meet a certain threshold/criteria. Even though the standards have defined the framework for interworking, the problem of network selection has been left out of the standard and has been the subject of a lot of research.

Several techniques have been proposed. These solutions have been realised either at the UE (distributed) or at the BS (centralised) and often aim to optimise a particular objective. Toni et al. proposed a decentralised network selection solution based on a multi-criteria utility function in [4] which aims to maximise the users quality of experience (QoE). Criteria such as monetary cost, network load, link quality, UE velocity and battery life are considered in the decision making process although no mention is made of how attainable all the parameters are or what happens in the absence of any. Moreover, since the network providing the highest utility is always selected for the user, this can lead to one network being 'popular', thus some networks could be heavily utilised while others could be under utilised. Multi Attribute Decision Making (MADM) based approaches to network selection have been proposed in [5] [6]. Similar to the utility based approach, these too have the tendency to choose the most 'popular' alternative thereby leading to load imbalance across the different networks. Hon et al. [7] model the network selection problem as a non-cooperative game amongst users wherein users try to maximise their throughput. The algorithm being centralised requires additional signalling between the BS and the UE, in particular to obtain information on neighbouring WiFi networks and to pass the network selection decision back to the UE. Decentralised game theory based approaches have also been proposed to address the aforementioned problem, e.g., [8] [9]. However, this shifts the computational burden on the UE which may be already constrained in terms of limited battery life unlike the BS. Also, game theory based approaches do not have an attractive convergence property. Stevens-Navarro et al [10] employ MDP to model the network selection problem and propose a decentralised algorithm to maximise the QoS for a UE's entire connection. At each periodic decision point, a mobile in its current state (using a network with certain bandwidth and delay conditions) would take an action to either remain in its current state, or move to a different state (a different network). The user would then be rewarded with a particular level of QoS. The reward function for moving from one state to the next is seen as a function of its rewarded bandwidth, delay and signalling cost. However, MDP is known to be computationally intensive and with an increase in the number of networks, states and actions, the complexity only increases. Another issue with MDP is that of the full execution of the algorithm to determine the best network at each periodic decision point, even when it is not warranted i.e. even if there is no change in state and it results in the same present network. Helou et al. [11] use utility theory to model the problem as a multi-criteria decision, in which the network providing the highest overall utility is selected. While they consider several applications, for instance streaming, elastic and inelastic applications, each with different QoS requirements and thus different utility requirements, they only consider the utility gained from throughput and monetary cost in the decision making. They do not consider other parameters (e.g. delay) that would be important to some applications.

In summary, a plethora of work has attempted to address the network selection problem using different techniques. Some of these optimise a limited set of parameters such as load or QoS but not both, some others are computationally expensive with the complexity increasing with scale whilst there are others that don't converge. What is desirable is a solution that optimises not only the QoS of the user but also balances the load across the networks and is simple from the perspective of a practical realisation. Such a potential solution is the subject of the ensuing discussion. *A key distinguishing feature of the proposed solution over the prior art is that it uses network attributes specific to the QoS class of the application in the network selection process.*

II. Proposed Network Selection Solution

In a loosely coupled architecture, the time taken for a cellular BS to receive updated information and make decisions may be too time consuming as there is no direct link between cellular BSs and WLAN APs to facilitate direct information exchange, and as such, this information will have to either be signalled via the UE or via the mobile core network, causing both increased signalling load and adding to delays. Given decisions are to be made at the mobile terminal, there is a requirement to conserve the devices battery life, thus providing a solution that is not too complex is absolutely paramount.

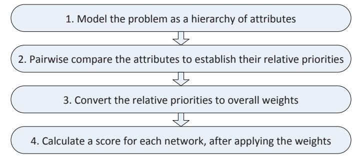


Fig. 1. Overview of the Analytic Hierarchy Process

TABLE I
PAIRWISE COMPARISON OF ATTRIBUTES (SOURCE: [12])

Relative Importance Level	e_{ij}	e_{ji}
i is equally as important as j	1	1
i is slightly more important than j	3	1/3
i is more important than j	5	1/5
i is much more important than j	7	1/7
i is extremely more important than j	9	1/9

Of the techniques mentioned in the previous section, both MADM and multi-criteria utility offer low computational complexity. However, there is a need to identify ways to improve loading across the RATs. Thus a method to improve resource utilisation is required. Additionally, the proposed algorithm aims to differentiate over the prior art by making use of network attributes that are specific to the QoS class of the data application. For instance, instead of simply using an average user throughput or delay value of a network in the decision making, it relies on the BS/AP to provide average throughput and delay values specific to real time, best effort class of users etc. In being a bit more granular with the attributes used, more precise decisions should be made, thus leading to enhanced QoS. Also, load values from the radio access technologies are incorporated into the decision making in an attempt to achieve better resource utilisation.

The proposed network selection algorithm is based on MADM, in which multiple criteria/attributes are considered in a structured way, to enable more informed decision making. A MADM problem can be solved by applying a well-known technique developed by Saaty (1980) [12] referred to as Analytical Hierarchy Process (AHP) (see Fig. 1).

Choice of attributes: Each network was assessed and ranked based on a set of attributes. The attributes considered were - 1) average throughput pertaining to the application

TABLE II
PREFERENCE MATRIX FOR REAL TIME VOICE APPLICATION

	Tput	Delay	Cost	Load	SINR
Tput	1	1/5	7	1	5
Delay	5	1	7	3	5
Cost	1/7	1/7	1	1/7	1/5
Load	1	1/3	7	1	3
SINR	1/5	1/5	5	1/3	1

TABLE III
PREFERENCE MATRIX FOR NON-REAL TIME BULK TRANSFER APPLICATION

	Tput	Delay	Cost	Load	SINR
Tput	1	7	5	1	3
Delay	1/7	1	1	1/7	1/5
Cost	1/5	1	1	1/7	1/5
Load	1	7	7	1	5
SINR	1/3	5	5	1/5	1

QoS class, 2) average delay pertaining to the application QoS class, 3) monetary cost, 4) SINR and 5) BS/AP load values. These attributes were chosen based on a reasonable assumption that they represent a fair set of parameters to form a basis for the decision making, but also due to the ability to realistically measure them in practice. Both throughput and delay are vital QoS parameters to be considered. Their values specific to the application QoS class were used based on the intuition that they would give a better indication of the anticipated performance of the network for that particular type of application. For instance, consider the case of a bulk transfer application for which throughput is the main QoS measure. A candidate network could advertise a high average throughput, but this could be due to real time video conferencing and real time gaming applications having high throughput with non-real time bulk transfer applications having low throughput. Thus using the average throughput on the network would not be a good indication of the typical experience for that type of application. However, using the average throughput specific to bulk transfer application on that network would be more appropriate. In addition to the main QoS parameters, monetary cost is also taken into account. Although a subscriber will be satisfied if his/her QoS is good, a subscriber would also be satisfied if his/her monetary cost is lower, thus monetary cost is included. Last but not least, in an attempt to achieve better resource utilisation across the networks, the BS/AP load values are considered in the decision making.

Pairwise comparison of attributes: This process involves comparing all the attributes against each other and assigning a value using a 9 point scale as shown in Table I where e_{ij} represents an element in a matrix of row i and column j. To ensure consistency when comparing two attributes, the rule $e_{ij} = 1/e_{ji}$ is enforced. The attributes throughput, delay, monetary cost, SINR and load are compared against each other to determine a preference matrix for the particular QoS class. For a real time voice service application, having minimal delay is more important than attaining high throughput. Also, for the load on the network to have an impact on the decision making, it is assumed to be equally as important as throughput. Thus, comparing the attributes led to the preference matrix for real time voice application as shown in table II. Similarly, for background non-real time services, for instance a file transfer application, throughput is more important than delay. Again, comparing the attributes for a non-real time bulk transfer service led to the preference matrix as shown in table III.

Translating attribute priorities into weights: The overall weightings for the attributes are then determined by using a method based on eigen vectors. By solving the equation $Ax = \lambda x$ for non-trivial solutions (i.e. $x \neq 0$), where A represents a square matrix (the preference matrix in this case) and represents eigenvalues, the corresponding solution to the equation or 'eigenvectors' for the preference matrix can be found. If the values of an eigenvector are normalised, they represent the overall weightings for the attributes, values of which add up to 1. The eigenvector solutions to the preference matrices for the real time and non real time applications were calculated in Matlab and were based upon the largest eigenvalue being used. The resulting normalised eigenvector, i.e., weighting for the real time voice service attributes and non-real time bulk transfer service attributes are shown in table IV.

TABLE IV
WEIGHTS OBTAINED FROM THE PREFERENCE MATRICES

	Tput	Delay	Cost	Load	SINR
Real time app.	0.2047	0.4952	0.0320	0.1871	0.081
Non-real time app.	0.3378	0.0441	0.0478	0.4154	0.155

TABLE V
DECISION MATRIX

	Tput	Delay	Cost	Load	SINR
WiFi	6.5Mbps	50ms	0	0.4	18dB
LTE	3Mbps	5ms	1	0.65	12dB

Scoring and ranking the networks: In this work, a heterogeneous network comprising LTE and WiFi radio access technologies was considered. Each network to be compared is represented by a row vector of attributes and is placed into a decision matrix as shown in a hypothetical example in table V. The network attribute values are then normalised to a common scale between 0 and 1, to enable comparisons across them to be made. Throughput and SINR are upward attributes, i.e., the higher the value, the more benefit brought to a subscriber. Thus, those attribute values were normalized against the highest value for that attribute across the networks being compared (normalised value = attribute value / max(attribute value across networks)). However, delay, monetary cost and AP/BS load are downward attributes. Thus, a higher value is actually worse for a subscriber and can be normalised as normalised value = (max(attribute value across networks) - attribute value)/max(attribute value across networks). For instance, a delay of 5ms on the LTE network, where the max delay was 50ms (on WiFi), the normalized LTE delay is (50ms-5ms)/50ms = 0.9. The normalised version of the decision matrix depicted in Table V is shown in Table VI.

In order to score each network, the overall weighting (according to the service application), is applied to the normalised decision matrix and the sum is taken. For instance, if the application was a non-real time bulk transfer application, when applied to the hypothetical normalised decision matrix, the score is determined by matrix multiplication (values in Table VI with the values corresponding to non real time application shown in Table IV to yield the scores of 0.700 for Wifi and 0.297 for LTE. The network with the highest score is then selected, in this case WiFi. To recap, the proposed network selection algorithm works as follows: *Get the attributes for networks within range of the UE, determine QoS class level when an application arrives, apply weights associated with this class to the network attributes to score and rank networks and finally choose the best network.*

TABLE VI
NORMALISED DECISION MATRIX

	Tput	Delay	Cost	Load	SINR
WiFi	1	0	1	0.384	1
LTE	0.46	0.9	0	0	0.66

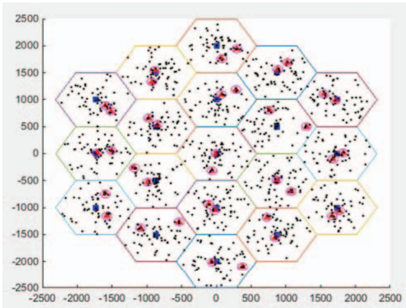


Fig. 2. Topology considered in the simulation study (axes unit - metres)

III. Performance Evaluation Methodology

A Monte Carlo 'snapshot based' system level simulator was developed in Matlab. A simulation took the form of multiple 'snapshots' of a model LTE/WLAN heterogeneous system in which UEs were randomly distributed in the simulation area and generating traffic according to statistical models. Only downlink performance was considered as typically traffic in the downlink is significantly higher than the uplink. The overall performance was then averaged across the multiple independent snapshots. The simulated network consisted of a LTE mobile network complemented by WLAN hotspots. 19 cells in a hexagonal arrangement were used to represent the coverage area of LTE macro base stations in a simulated environment, in accordance with 3GPP TR36.942 [13]. ENodeBs, each with a height of 30m, were placed in fixed locations at the centre of the hexagonal cell with an inter-site distance (ISD) of 1000 metres. A fixed number of Wi-Fi Access Points, with a fixed height of 3m, were then placed at random co-ordinates within each LTE BS area. In initial simulations, 2 Wi-Fi APs per BS cell were used. However, this was a variable parameter in the simulation. 60 UEs, with a height of 1.5m were then distributed randomly, according to a uniform distribution within each hexagonal cell, with a fraction of UEs being located within the range of Wi-Fi APs. A fraction of 8/15 was used, however, this was also a parameter that was varied in the simulation. The Cartesian (x, y, z) co-ordinates of all LTE BSs, Wi-Fi APs and UEs within the simulation environment were known. Given this information, the separation distances and thus propagation losses due to distance were then calculated. Fig. 2 shows the topology and Table VII and VIII highlight the values of the simulation parameters chosen based on the 3GPP and IEEE 802.11 simulation scenarios captured in [13] and [14].

Real time voice and non real time bulk transfer traffic was considered in the simulations. The real time voice session was characterised by a source generating packets at a rate of 12.2kbps with a frame size of 36bytes. The non real time bulk transfer source was characterised by a 0.5MB file download with a frame size of 1500bytes. Session arrival rate for both the traffic types was assumed to be Poisson distributed with mean arrival time of λ which was varied from 0.5 to 2.5 (in steps of 0.5) during the simulation. A round robin scheduler was employed in the case of both LTE and WiFi. Given the large frame size for the bulk transfer application, RTS-CTS was enabled. On the other hand this was disabled for the voice application due to the small frame size. Following four algorithms were considered in the simulation study:

TABLE VII
LTE SIMULATION PARAMETERS

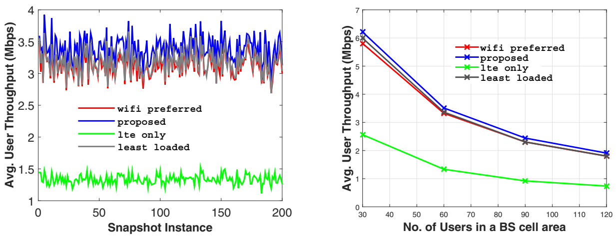
Environment	Urban
Cellular layout	19 BSs in a hexagonal grid with the base station at the centre of the cell (omni-directional antennas)
Carrier Frequency	2 GHz (3GPPs simulations for EUTRA are based on 2GHz)
BS Antenna Gain	15 dBi
UE Antenna Gain	0 dBi
BS Antenna Height	30 m
Inter-site Distance	1000 m
Path-loss Model	128.1 + 37.6 log10 (R) + χ (where R represents the UE/BS separation distance in km, χ represents the log normal shadowing term)
Log Normal Shadow Variance	10 dB
Minimum Coupling Loss	70 dB
White Noise Power Density	-174 dBm/Hz
System Bandwidth	10 MHz FDD
BS Max Tx Power	46 dBm
UE Noise Figure	9dB
Traffic Model	Finite Buffer
Resource Block Size	180kHz

TABLE VIII
WIFI SIMULATION PARAMETERS

AP Tx Power	20 dBm
Carrier Frequency	5 GHz
System Bandwidth	20 MHz
AP Antenna Gain	0 dBi
AP Antenna Height	3 m
Distance based Pathloss	For distances d<10m (PL = 40.05 + 20 log10(fc/2.4) + 20 log10(d) + χ), For distances d>10m (PL = 40.05 + 20 log10(fc/2.4) + 20 log10(d) + 35 log10(d/10) + χ) where and fc is the carrier frequency in GHz
Shadowing	5 dB standard deviation

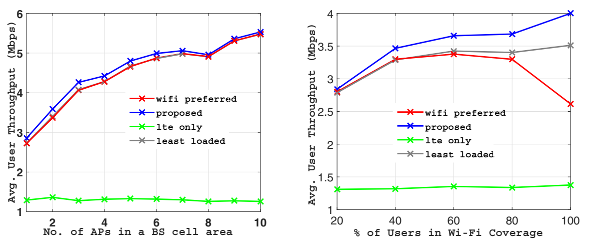
- **Wifi Preferred:** UE selects WiFi whenever it is within WiFi coverage.
- **Least Loaded Network:** UE selects the least loaded network
- **LTE Only:** UE always selects LTE even though it may be in WiFi coverage.
- **Proposed Algorithm:** UE ranks networks and chooses the best as per criteria explained earlier.

In general, a simulation consisted of 200 snapshots in order to cover a wide range of UE locations within the simulation area. In each snapshot, 60 UEs were placed randomly within each BS cell area and the traffic generated within a 5 second period (i.e. 5000 subframes) was simulated. For the baseline scenario, there were 2 APs in each BS cell area, roughly half the number of users within dual coverage and UEs generating new sessions at an arrival rate of 0.5. The effect of varying the different parameters within the baseline scenario were investigated and the performance of the algorithms was then evaluated. The performance metrics considered in this study were the average user throughput, medium access queuing delay and the average system throughput.



(a) Baseline scenario (b) Varying no. of users

Fig. 3. Average user throughput



(a) Varying no. of APs per BS (b) % Users in WiFi coverage

Fig. 4. Average user throughput

IV. Results

In general, the proposed algorithm was shown to provide a higher average throughput for users compared to the other algorithms under the baseline simulation scenario across all 200 snapshots as evident from Fig. 3(a). Even when the number of users per BS cell area was varied, the proposed algorithm provided a higher average user throughput (Fig. 3(b)). However, as the number of users in the cell area increases to 120, the average user throughput provided by the proposed algorithm began to converge to that of other conventional algorithms. This is expected, since under increasing load, the system starts to become saturated and there are limited ways in which the capacity can be shared across the users.

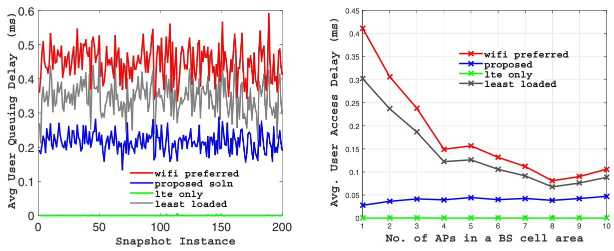
Fig. 4(a) shows the average throughput for a scenario wherein the number of APs within each cell area was varied between 1 and 10, while keeping other parameters in the baseline scenario constant. The average user throughput provided by the proposed algorithm was seen to be higher, although this was seen to converge to that of the other algorithms with increasing number of APs (8 or more APs). Bearing in mind that the number of users within each BS cell area is fixed at 60 and roughly half of the users are in Wi-Fi (dual) coverage, as the number of APs increase, roughly 30 users are split across the APs. With less users in Wi-Fi (dual) coverage, it is highly likely that Wi-Fi would be the less loaded of the two networks and that Wi-Fi would offer better performance compared to LTE and thus be selected. Thus the proposed algorithm provides similar performance to the main conventional algorithms under extremely low load situations as is the case here with several APs, each with a few users.

Although it is not anticipated for there to be a drop in average user throughput with increasing number of APs, the dip in user throughput seen when there are 4 or 8 APs, could be due to the random distribution of the locations of the APs within the BS cell area. If APs are located in close proximity to each other, the interference experienced by a UE would increase and this would result in both reduced SINR values, reduced achievable PHY data rates and hence lower achievable throughput for the UEs. Also, the average user throughput obtained by using the 'LTE only' algorithm should be roughly the same, regardless of the number of APs within the cell area. The slight variation seen in average user throughput is simply due to the random nature of the simulation, since users have random locations in each snapshot, and by extension, in each simulation.

Another interesting observation is the similar performance seen between the 'Wi-Fi preferred' algorithm and the 'least loaded' algorithm, especially as the number of APs increase (evident by the overlap in their plots). This seems to suggest that the Wi-Fi network was the lesser loaded of the two networks and was selected by the least loaded algorithm. This illustrates the point that while the load metrics are able to provide a view of load on their respective radio networks, they do not necessarily give an indication of whether there is more capacity available in order to achieve a higher throughput. A higher user throughput can be achieved by including other metrics than simply the load, as evident with the proposed algorithm.

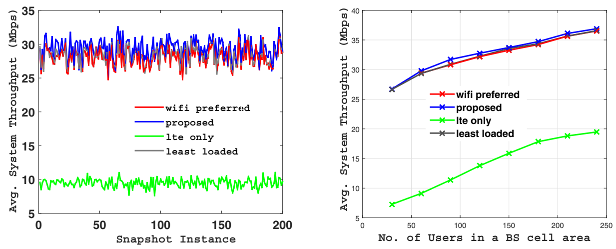
Fig. 4(b) shows the average throughput for a scenario wherein the percentage of UEs within Wi-Fi coverage was varied incrementally between 20-100%. Observe that the proposed algorithm provides higher average user throughput except for the case in which 20% of users were within Wi-Fi AP coverage where comparable performance was seen. In that case, 6 of the 60 users were in each of the 2 Wi-Fi AP coverage areas and again there was extremely low load on the APs since not all users would be active simultaneously. Thus, the Wi-Fi network may be less loaded and may be selected in preference to LTE by all the main algorithms and thus similar performance would be seen. It is also interesting to note the vast difference in performance between the algorithms when all 60 LTE BS cell users (i.e. 100% of users) are within Wi-Fi (dual) coverage. The proposed algorithm is able to provide much higher throughput for users. While the 'Wi-Fi preferred' algorithm is only limited to using the capacity available from the Wi-Fi network amongst users (thus achieving lower average throughput), the 'least loaded' and proposed algorithms are able to use the capacity provided by both the LTE and Wi-Fi networks. Thus higher average throughput is achieved by the 'least loaded' and proposed algorithm. However, the proposed algorithm still outperforms the 'least loaded' algorithm.

We now turn our attention to the medium access queuing delay. Fig. 5(a) shows the queuing delay (at the MAC layer) experienced by the users for the different algorithms for the baseline scenario. In general, the proposed algorithm provided lower queuing delays than the conventional algorithms. It is important to note that for the LTE network, since the processing delays are ignored, if a packet is generated at the MAC layer at a particular time instant (subframe time), the packet can be scheduled during that subframe, once there are RBs available. Hence the queuing delay would be 0. In all sim-



(a) Baseline scenario (b) Varying no. of APs

Fig. 5. Average queuing delay



(a) Baseline scenario (b) Varying no. of APs

Fig. 6. Average system throughput

ulated scenarios, the proposed algorithm was able to provide low queuing delays compared to the other algorithms. One of the interesting findings (Fig. 5(b)) was that the proposed algorithm can provide significantly lower delays compared to the conventional algorithms when there are only a few APs in the system. When the number of APs increase, the load per AP reduces and hence all algorithms achieve a low delay. The proposed algorithm however still outperforms the other two.

The average system throughput provides an indication of how well the algorithms are making use of the available resources. In the simulation, it is measured by summing the average throughput of a LTE BS cell and the average throughput of an AP within the simulation area. The system throughput depends on the offered load provided to the networks. Although there are ultimately limitations to the max achievable LTE cell throughput in the system since only a round robin scheduler is employed which does not seek to maximise throughput, the performance achieved while using the different algorithms can still be compared against each other. For the baseline simulation across 200 snapshots, the proposed solution provided slightly higher system throughput in comparison to the other algorithms as seen from Fig. 6(a). Even when the number of users within the simulation area is increased, keeping other parameters constant, the proposed solution provided slightly better system throughput as evident from Fig. 6(b). As expected though, this performance converges to that of the other algorithms as the number of users increase. The proposed algorithm was also found to outperform the other algorithms in the scenario where the session arrival rate was increased. These results have however been left out due to lack of space.

V. Conclusion

Having an algorithm that improves both QoS and resource utilisation is important in heterogeneous mobile networks especially as operators deploy Voice over Wi-Fi, where regular Voice data services have to compete with other data traffic on a contention based network. In this paper, a network selection algorithm taking into account class based QoS performance advertised by the network was presented and was shown to provide a good trade-off between complexity, QoS provision and improving system performance. The proposed algorithm was found to provide enhanced QoS for users compared to conventional algorithms, such as a 'WiFi preferred', by means of higher throughput and lower delay tailored to the QoS class of the user. For instance, it exhibited the ability to offer non real time users higher throughput and real time users lower delay. The algorithm was also seen to provide significant performance gains when there was a higher proportion of users within coverage of more than one network. Looking forward, future work could investigate enhancements to the algorithm to include other QoS parameters such as PER and PLR, or periodic execution during run time so as to assess handover performance. Also the simulation model could benefit by modelling mobility or including more QoS classes such as video and web traffic.

References

[1] Ericsson, "Wifi in heterogeneous networks," http://www.ericsson.com/res/docs/whitepapers/wp-wi-fi-in-heterogeneous-networks.pdf, Sep 2015.

[2] R. Wireless, "A detailed look at hotspot 2.0," http://www.ruckuswireless.com/technology/hotspot2, Sep 2015.

[3] 3GPP, "Access network discovery and selection function (andsf) management object (mo), ts 24.312 v12.7.0," http://www.3gpp.org/dynareport/24312.htm, Dec 2014.

[4] N. Quoc-Thinh, N. Agoulmine, E. Cherkaoui, and L. Toni, "Multicriteria Optimization of Access Selection to Improve the Quality of Experience in Heterogeneous Wireless Access Networks," *IEEE Transactions on Vehicular Technology*, vol. 62, pp. 1785–1800, 2013.

[5] C. Jin, W. Zaixue, W. Yu, S. Lin, and Y. Dacheng, "A service-adaptive multi-criteria vertical handoff algorithm in heterogeneous wireless networks," in *IEEE PIMRC*, Sep 2012.

[6] S. Maaloul, M. Afif, and S. Tabbane, "Vertical Handover Decision Policy Based on the End User's Perceived Quality of Service," in *IEEE AINA Workshops*, Mar 2013.

[7] C. M. Hon, R. Southwell, and H. Jianwei, "Congestion-aware network selection and data offloading," in *48th Annual Conference on Information Sciences and Systems (CISS)*, 2014.

[8] T. Li-Chuan, C. Feng-Tsun, Z. Daqiang, R. Y. Chang, C. Wei-Ho, and H. ChingYao, "Network Selection in Cognitive Heterogeneous Networks Using Stochastic Learning," *IEEE Communications Letters*, vol. 17, pp. 2304–2307, 2013.

[9] Z. Kun, D. Niyato, and W. Ping, "Network Selection in Heterogeneous Wireless Networks: Evolution with Incomplete Information," in *IEEE WCNC*, 2010.

[10] E. Stevens-Navarro, L. Yuxia, and V. W. S. Wong, "An MDP-Based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 57, pp. 1243–1254, 2008.

[11] M. E. Helou, S. Lahoud, M. Ibrahim, and K. Khawam, "A Hybrid Approach for Radio Access Technology Selection in Heterogeneous Wireless Networks," in *19th European Wireless Conference*, 2013.

[12] S. French, *Decision theory : an introduction to the mathematics of rationality*. Chichester: Ellis Horwood, 1988.

[13] 3GPP, "E-utra radio frequency system scenarios (release 12), ts 36.942 v12.0.0," http://www.3gpp.org/dynareport/36942.htm, Sep 2014.

[14] S. Merlin et al., "IEEE 802.11-14/0980r12 TGax Simulation Scenarios," IEEE, Tech. Rep., May 2015.