Probabilistic multi-RAT performance abstractions

Akhila Rao and Rebecca Steinert
RISE SICS AB
{akhila.rao, rebecca.steinert}@ri.se

Abstract—Development towards 5G has introduced difficult challenges in effectively managing and operating heterogeneous infrastructures under highly varying network conditions. Enabling, for example, unified coordination and management of radio resources across coexisting, multiple radio access technologies (multi-RAT), require efficient representation using high-level abstractions of the radio network performance and state. Without such abstractions, users and networks cannot harvest the full potential of increased resource density and connectivity options resulting in failure to meet the ambitions of 5G.

We present a generic probabilistic approach for unified estimation of performance variability based on attainable throughput of UDP traffic in multi-RATs, and evaluate the applicability in an interface selection control case (involving WiFi and LTE) based on obtaining probabilistic user performance guarantees. From simulations we observe that both users and operators can significantly benefit from this improved service availability at low network cost. Initial results indicate 1) 116% fewer performance violations and 2) 20% fewer performance violations with a reduction by 35 times in the number of handovers, compared to naive and state-of-the-art baselines, respectively.

Index Terms—probabilistic modelling; heterogeneous networks; multi-RAT networks; interface selection

I. INTRODUCTION

High user throughput, predictable performance, efficient resource utilization and coordination are some of the objectives of 5G [1], [2]. Dense heterogeneous networks under 5G are characterized by a dramatic increase in the number of mobile nodes, mobility over smaller coverage areas, dynamically changing resource availability (duty cycling base stations), interference, competing standards, multiple connectivity options, etc [3]. Under such conditions, effectively dealing with network and service performance at scale, quickly becomes a technically overwhelming and expensive task for the operators.

To achieve the technical objectives of 5G [2] and create networks that are manageable, controllable and operable across heterogeneous infrastructures, the telecom industry and the research community emphasize the need of developing intelligent mechanisms for network management and control. In this context, a fundamental component is the ability to observe and coordinate network resources across heterogeneous RATs using unified high-level abstractions [4]. Development of such monitoring and control methods continue to be a challenge, due to 1) scalability limitations in making direct measurements and, 2) observability limitations in passive observations of related metrics. Using existing metrics in 5G radio networks, such as SINR [5], is not sufficient for effective resource and performance management, as such metrics fail to capture

necessary aspects of resource demands or availability. Instead, intelligent monitoring approaches combining a broader set of available service and network metrics, thereby providing richer information suitable for automated management decisions across heterogeneous infrastructures, are essential in achieving 5G performance objectives.

We propose *AQuaMet* (*A*gnostic *Qua*lity *Met*ric), a probabilistic abstraction of the network state based on downlink *attainable throughput* that quantifies performance variability in the radio network and enables informed decision making for policy enforcement through coordinated control. AQuaMet uses analytical models from prior work in [6] and [7] and builds a predictive model that combines current and historical observations to provide fast, yet robust, probabilistic estimates of downlink attainable throughput. Probabilistic attainable throughput offers a comparable way of estimating link performance across multi-RATs, by comparing the probability of violating QoS requirements at the user. The approach is applicable to scenarios where a service provider needs to ensure a subscribed level of throughput to a user in a dynamic multi-RAT network, specified by probabilistic guarantees. An example application requiring such robustness is high quality video conferencing (e.g. Skype), which needs a consistent level of throughput to maintain a satisfactory frame rate [8].

Further, we demonstrate the applicability of AQuaMet supporting a controller application for downlink interface selection in a multi-RAT scenario involving WiFi and LTE. Carrier companies currently explore the extension of licensed LTE spectrum with unlicensed spectrum available to WiFi in scenarios involving LTE small-cell and WiFi AP deployments. We thus motivate our choice of the most relevant heterogeneous network scenario where WiFi and LTE networks coexist under a common controller. Note that, although the evaluation addresses a specific WiFi-LTE scenario, the use of AQuaMet in management decisions can be generalized to any RAT and also extended to include other QoS requirements, such as delay.

AQuaMet is a computationally light-weight approach built on pre-existing metrics and is compliant with current standards, which makes it practically applicable in real-world networks along with great implementation flexibility (decentralized or centralized). Initial results from synthetically generated scenarios in NS3 indicate that using AQuaMet for managing resources with downlink interface selection, reduces the number of handovers by 35 times along with reducing performance violations by 20% and 116%, compared to the state-of-the-art [6] and naive baselines (still under use in e.g. enterprise WiFi) respectively.

A. Related work

Control and management applications in multi-RAT networks require RAT-agnostic representations that can be used to compare user performance across multiple RATs. SINR has previously been explored as a metric for dynamic RAT selection [5]. However, SINR is not an effective metric in capturing user performance resulting in development of load and throughput based metrics. In [9] and [10], the authors use a simplistic estimation of throughput from SINR and bandwidth using the Shannon capacity equation. The authors of [11] and [12] use a normalized load metric over multiple RATs as an indicator of performance. Deng et al. [13] use active probing to characterize links and select between available RATs for each user. The approach in [6] uses average attainable throughput over time windows of seconds to minutes, which is insufficient to accurately capture radio performance variability. Yan et al. [14] and Chakraborthy et al. [15] apply learning methods for performance estimation, which require large amounts of data and long training times.

In most real-world networks the current practice is still to use thresholds on only low-level radio parameters such as RSSI, SINR or network load for controller applications such as interface selection, load balancing etc [5][11]. These parameters are not informative in describing the true user performance from a user application's perspective. Even approaches that have used more high-level metrics such as attainable throughput [6] average over variations in the network resulting in less informative representations. Other methods that use active probe traffic over available links have severe scalability restrictions, due to high measurement overhead in large networks. Learning methods, although promising, are not currently light-weight enough to adapt quickly to fast changes in the network state and can take hours to train in new environments [16]. These approaches from prior work have not been designed with the focused objective of meeting the challenges of highly dynamic multi-RAT networks which will be the norm in 5G [2].

B. Contribution

We address the aforementioned shortcomings of prior work and advance the state-of-the-art of representing link performance in terms of RAT-agnostic high-level abstractions by:

- Detailing a monitoring mechanism that is designed to be computationally light-weight, scalable and practically applicable in real-world radio networks.
- Formulating a generic probabilistic estimator for attainable throughput based on combining existing available metrics on top of [6] and [7].
- Evaluating the applicability of the approach in one controller application case focused on interface selection, relative to baseline implementations [5] [6].

II. OVERVIEW OF THE APPROACH

The implementation of AQuaMet consists of distributed monitoring instances at each *serving node* (AP or base station) and a monitoring/management instance at a logically
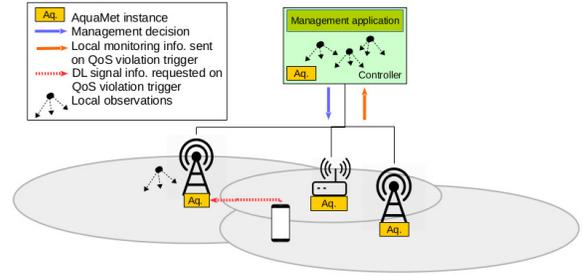

Fig. 1: Multi-RAT heterogeneous network with common coordination and control using AQuaMet.

centralised controller located at the network gateway as shown in Figure 1. In this context, the envisioned application of AQuaMet is to provide controller support for the dynamic monitoring of link and network performance for interface selection with the objective of meeting QoS guarantees. AQuaMet can also be used for other control applications relying on probabilistic attainable throughput (e.g. load balancing).

The AQuaMet monitoring instance at each serving node makes local observations and sends monitoring information to the controller only when triggered due to a QoS violation on a user. The controller aggregates local observations from candidate serving nodes and estimates probabilistic attainable throughput on candidate links in the network, to make an informed coordinated management decision. This makes the approach computationally light-weight overcoming the need for periodic estimation of attainable throughput on every candidate link in the network. We make passive observations, accumulating relevant base metrics over short measurement time windows W_m. The metrics accumulated over each W_m are aggregated in a sliding window W_s. The size of the sliding window is based on the time scale at which stable QoS has been guaranteed (with a probabilistic threshold), and is dependent on the requirements of user applications. The metrics for downlink traffic related observations are available at the serving nodes, and metrics for downlink signal quality observation are pulled from user nodes or *clients* only when a management action is triggered.

An ECDF over measured downlink throughput of clients at each serving node monitors the probability of violating its probabilistic performance requirement. When this probability exceeds a tolerance level (detailed in Section IV), the serving node informs the controller, which triggers a management action. When one or more clients trigger the management function, the base metrics aggregated over W_s are communicated to the controller so that the management function has a full network view encompassing all available candidate links. Attainable throughput is estimated at the controller using analytical models from prior work in [6] and [7] using the base metrics in each W_m. For each available candidate link, this is obtained in the form of ECDF of the attainable throughput.

Modelling using known parametric distributions require fewer samples and are computationally more efficient in prediction since it does not scale with the number of samples. However, the scenario of 5G networks involves varying traffic patterns, access patterns, and network conditions, wherein, using a tailored parametric distribution would limit the scope of

monitoring and introduce errors when the network conditions shifts between different regimes. Our approach of using a non-parametric representation of performance on a link using ECDFs, provides a solution that can be generalized over the aforementioned variations since the distributions do not have to fit a parametric representation, chosen a-priori.

Current commercial wireless network implementations use downlink information to perform mobility management actions, since most traffic in commercial wireless networks is downlink. Also, emergence of high throughput UDP based transport solutions, such as random linear network coding [17] motivate our focus on downlink UDP throughput. In the following section we will detail the analytical model used for estimation of attainable throughput on candidate links.

III. ESTIMATING ATTAINABLE THROUGHPUT

Models from prior work in [6] and [7], with input base metrics is used for the estimation of attainable throughput in both WiFi and LTE RATs. The choice of these models was based on their ability to estimate the attainable throughput for any potential client-serving node association sets. This allows the switching of multiple clients to new target serving nodes at the same evaluation instant. The base metrics used in the estimation model are presented below.

A. Metrics

1) WiFi: **RSSI** Received signal strength indicator is a metric for relative power of received signal. Downlink RSSI at the client is communicated uplink to the serving node by means of measurement reports provided by the 802.11k [18]. **Packet delivery ratio (pdr)** captures losses resulting from fading as well as collision. **Arrival rate (arr)** in packets per unit time and average **Packet length (p)** is measured for each user. **Number of active users** (N_a) represents the number of users accessing network resources in the current measurement time window. **Signal to noise ratio** (SNR) is computed from RSSI and noise at the receiver, and informs about the modulation and coding scheme (MCS) setting and sending rate (m) at the transmitter, based on target thresholds for bit error rates (BER) [19].

2) LTE: **RSRP** Reference signal received power is a measure of received signal strength which is communicated uplink to the serving node using standard specified measurement reports. **SINR** is obtained from RSRP along with interference and noise at the receiver. SINR (represented as channel quality indicator-CQI), informs the setting of MCS and sending rate (m) at the transmitter using target thresholds for transport block error ratio (TBLER) [20]. The arrival rate, packet length, and number of active user measurements are common for both LTE and WiFi.

B. Estimation

1) WiFi: In WiFi networks, the wireless spectrum is shared between users by means of random access using CSMA/CA. In downlink, the AP MAC performs first-in-first-out over the packets in its input queue. We use the model for saturated WiFi

throughput in [6] combined with an estimate for unsaturated network scenario, and summarize the equations in (1), (2), and (3). Here, k corresponds to measurement time windows and i corresponds to the users in the network. The estimated attainable throughput λ_k^i in a measurement time window k is evaluated for each client i. T_{ack} is the time to send a MAC acknowledgement using the appropriate m_{ack} as obtained from the data MCS m. T_{phy} is the time to send the physical layer header. $DIFS$ and $SIFS$ are standard specified constants.

$$\lambda_{sat,k}^i = \frac{pdr_k^i \times arr_k^i \times p_k^i}{\sum_i arr_k^i \left(\frac{p_k^i}{m_k^i} + DIFS + SIFS + T_{ack,k}^i + T_{phy} \right)} \tag{1}$$

$$\lambda_{unsat,k}^i = arr_k^i \times p_k^i \times pdr_k^i \tag{2}$$

$$\lambda_k^i = min \left(\lambda_{sat,k}^i, \lambda_{unsat,k}^i \right) \tag{3}$$

2) LTE: On the other hand, LTE networks have scheduled access to spectrum resources. The downlink throughput at the client in an LTE network depends on the MAC scheduler at the eNodeB used to assign resources, and its fairness objectives. The proportional fair (PF) scheduler provides throughput fairness among the clients in the network [21]. In practice, most comparative experiments on LTE networks are performed using PF since it is a widely used scheduler. While there are other schedulers, we employ PF scheduling in line with common practice [7]. The throughput estimation model is specific to PF, but can be extended for other schedulers based on their resource allocation objectives.

We use prior work presented in [7] to estimate the share of LTE bandwidth resources (Y_k^i) allocated to each user. As described in [7], this allocation is based on the amount of requested resources and the current load conditions. The number of bits that can be sent in each scheduling instance, TBS, is a function of the MCS m_k^i on the link and the data bandwidth available at the sending eNodeB. The expression for attainable throughput, obtained from [7] is given by:

$$\lambda_k^i = min \left(Y_k^i \times TBS_k^i, Y_k^i \times p_k^i \right) \tag{4}$$

IV. PROBABILISTIC INTERFACE SELECTION
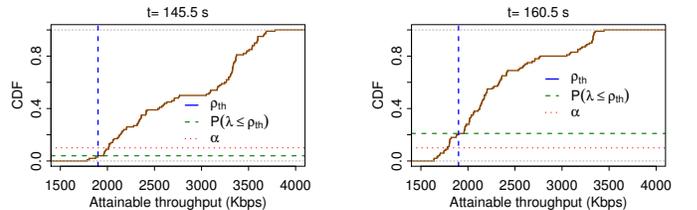
(a) Performance objective met.　　(b) Performance objective not met.

Fig. 2: Evaluating throughput objective on candidate links.

The probabilistic QoS, as given by expression (5) on a link to the serving node (serving link), is obtained using direct observation of throughput over W_m. Here, ρ is the observed throughput, ρ_{th} is the required threshold and α is the tolerance to this throughput being less than ρ_{th}. Probabilistic QoS on a *candidate link*, as estimated from equations (3) and (4) is given by the expression (6), where λ is the attainable throughput:

$$P\left(\rho \leq \rho_{th}\right) \leq \alpha \qquad (5)$$

$$P\left(\lambda \leq \rho_{th}\right) \leq \alpha \qquad (6)$$

Given a tolerance α of, say 10%, we see in the example in Figure 2 how the conditions shift from $P\left(\lambda \leq \rho_{th}\right) \leq \alpha$ towards breaching the tolerance level with $P\left(\lambda \leq \rho_{th}\right) > \alpha$ in a time gap of 15 s. When a serving link fails to satisfy the condition given by expression (5), the controller is triggered to select a candidate link that satisfies expression (6), if available. It is to be noted that in some scenarios no available candidate link will satisfy this requirement, and the link with the least deviation from the requirement is chosen. Robustness is added to the decisions made by AQuaMet thanks to its sliding window approach that weights historical observations along with current observations in the time scale of interest.

The design of AQuaMet is light-weight - both computationally and in terms of monitoring overhead. The memory requirement at the local AQuaMet instance on a serving node and at the controller scale linearly with the size of W_s, the number of clients served and number of serving nodes. The bandwidth required to pull signal strength information from clients also scales with W_s and the number of clients, but is still modest since it is only pulled on the event of a QoS violation trigger. Therefore, AQuaMet enables significant overhead reduction by using passive monitoring and with monitoring data sent to the controller only upon performance degradation.

V. EVALUATION SET-UP AND RESULTS

In this section, we verify the attainable throughput estimation models for WiFi and LTE from prior work [6] [7] and its applicability to be used in a unified abstraction. We then evaluate the use of AQuaMet abstractions for interface selection. All evaluations are performed on the NS3 simulator [22].

A. Estimation model

The objective of evaluating the estimation model is to observe its accuracy over candidate links in dynamic conditions. However, to be able to evaluate its accuracy, we need to know the ground truth. Hence, we estimate attainable throughput using the same metrics over serving links and use the measured throughput as ground truth. We also study the impact of the network size and the time over which metrics are aggregated on estimation accuracy.

Setup: The simulation setup consists of a network with a serving node (eNodeB or AP) and a varying number of mobile clients associated with this serving node. The experiments were done separately for WiFi and LTE. The clients are initialized at uniform random locations and follow a random waypoint mobility model. Downlink UDP traffic is injected from a remote host to the clients, with randomized start times and inter packet delays. The input data rate for each client is chosen randomly from a uniform set, representative of video streaming data rates (1.5 Mbps-8 Mbps) [23]. Attainable
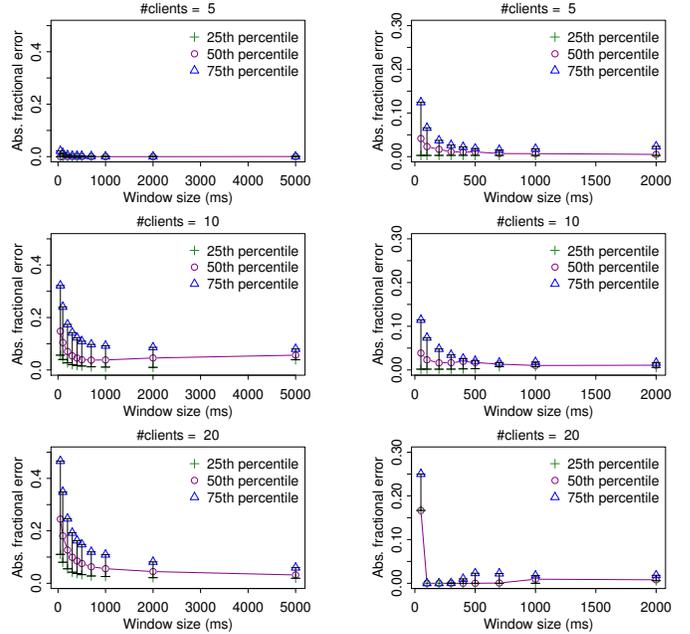

Fig. 3: Estimation error for WiFi (a,c,e) is below 10% using a window size of 500 ms or higher and for LTE (b,d,f) is below 5% using a window size of 500 ms or higher, despite increasing number of users (and thereby load). Because of scheduled access to resource in LTE, λ is more predictable than for WiFi.

throughput is estimated using base metrics in each W_m, based on the model presented in Section III. This estimate λ is compared against the ground truth throughput ρ to obtain estimation errors as $e = \frac{|\lambda - \rho|}{\rho}$. The experiments for both WiFi and LTE are repeated 20 times for statistical significance.

Results: We first present the results for the WiFi RAT, followed by the results for LTE RAT. Figure 3 (a,c,e) shows the absolute fractional error of estimated attainable throughput, versus increasing W_m, where 25th, 50th, and 75th percentiles of the error have been plotted. We see that the estimation error e is low for smaller number of clients, when the network is more likely to be in an unsaturated state. Moreover, we see that when increasing the number of nodes up to 10 clients, the estimation error rate increases, but remains about the same at 20 clients. This is because, when the network approaches saturation a large number of clients are competing for radio resources resulting in higher variance, with its degree given by the time scale of observation W_m. When W_m is increased, e reduces, because the number of observable packets is higher, providing more accurate estimation. This graph can be used as a tool to select W_m large enough to estimate throughput with a sufficient accuracy.

Moving on to LTE, resource access is scheduled and allocated to clients based on their requirement and a fairness objective. This is a point of contrast between LTE and WiFi, which has random access. We use PF scheduler in our approach as justified in Section III-B2. The simulation experiments performed for LTE were the same as for the WiFi RAT. Figure 3 (b,d,f) depicts the absolute fractional error as a function of W_m for varying number of clients. Throughput estimation error reduces with increasing W_m, and

drops faster than observed with WiFi in Figure 3 (a,c,e). Estimation accuracy close to ground truth is achieved by fine-tuning the parameters of the method, using graphs shown in Figure 3 as guidelines. To conclude, these results verify the applicability of the throughput estimation model.

B. Probabilistic interface selection

AQuaMet is generally applicable to QoS based control applications - here we evaluate its performance in the case of interface selection relative to three baseline methods. **SS**: A naive approach based on measured received signal strength in dBm. Interface selection is done based on highest received signal strength. **ATOM**: Estimates average attainable throughput in 10 s windows and selects interface with the highest value [6]. **IATP**: Estimate instantaneous attainable throughput using ATOM over small windows of 500 ms and select the interface with the highest value.

The first baseline was chosen to make a comparison with a simple low-level metric. ATOM was chosen as a baseline since, to the best of our knowledge, it is the only evaluation of interface selection in a multi-RAT network that evaluates attainable throughput [6] on candidate links while also considering its impact on other clients at the same serving node. In the ATOM baseline, attainable throughput is obtained as an average over a time duration of 10 s, as used in the paper. It is called average throughput since it attempts to average over the variations in the network state. The IATP baseline, which we defined by derivation from ATOM method, is an estimation over very small time windows which are comparable to the true rate of underlying network state variations.

Setup: The interface selection evaluation setup consists of one WiFi and one LTE serving node, with 20 mobile clients in the network. A remote host generates downlink traffic towards the clients. Interface selection is performed by updating the serving node to client association set for a target client node. In our simulation the interface selection and update of association sets are done virtually without performing an actual vertical handover. The W_m for AQuaMet is chosen to be 500 ms based on our observation of accuracy versus W_m from our results for estimation of attainable throughput presented in Section V-A. The value of W_s is 50 s with 100 samples of W_m in each window. This captures historical variations from the past 50 s on the link. The estimation window was chosen to be 10 s for ATOM (as in [6]), 10 s for SS, and 500 ms for IATP. The throughput threshold (ρ_{th}) was set at 2000 Kbps based on the range of input traffic rates. The QoS violation tolerance level (α) was set at 10%.

Results: Figure 4 shows the results of the comparison of AQuaMet with the previously described baselines. Figure 4a compares the ECDFs of number of performance violations (i.e. number of instances where the condition $P(\lambda \leq \rho_{th}) > \alpha$ is true). We reiterate that at times none of the available serving nodes can meet the requirements of some users, in which case the link with the best performance is selected. We see that AQuaMet performs better than SS and ATOM, and has relatively the same performance as IATP.
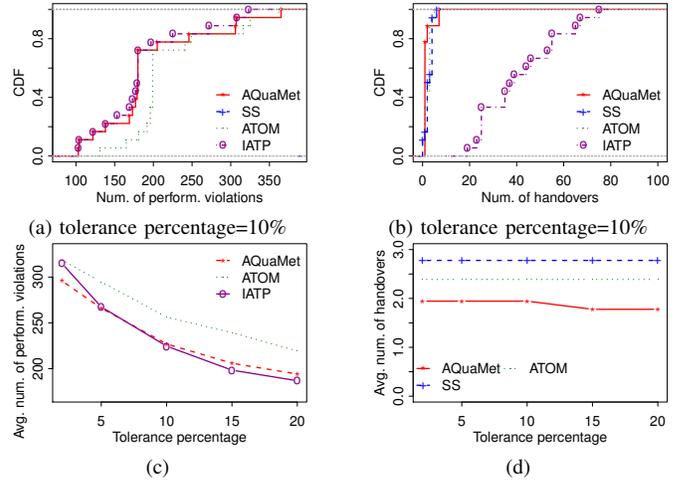

Fig. 4: (a,c) AQuaMet results in fewer average number of performance violations in comparison with ATOM. Number of performance violations for SS is in the range 558-589 and hence omitted for clarity. (b,d) Average number of handovers, with AQuaMet, decreases with increasing tolerance percentage. Average number of handovers for IATP is 41.6, which is much larger than others and hence omitted.

Figure 4b shows the number of interface changes or handovers performed over the same experiments. We see that, while the number of throughput performance violation using IATP is comparably similar to AQuaMet, it comes at the cost of a significantly higher number of interface changes, which can be prohibitive for large networks. Since IATP opportunistically switches between interfaces at intervals as small as 500 ms, it follows the changes in the network to achieve a performance similar to AQuaMet. This however comes at the cost of 35 times more handovers which leads to high unnecessary cost at the expense of resource efficiency and service availability. Thanks to the robustness of AQuaMet's probabilistic modelling, it achieves the same performance as IATP, but with fewer handovers (Figure 4b).

Figure 4c depicts the average number of performance violations at increasing tolerance levels (α). We see that when evaluating at any tolerance level in the range 2%-20%, AQuaMet has fewer violations than ATOM and is close to the performance of IATP. The number of performance violations for SS baseline is over 3 times higher than for the other methods and is hence omitted from the graph for clarity of observation of better methods. The gain in throughput performance compared to SS is because, unlike SS, AQuaMet interface selection decisions consider both signal quality and network load. This is also the reason the number of performance violations for SS is very large. The gain compared to ATOM is from using a probabilistic metric that triggers interface selection in response to variation in the underlying distribution.

The result in Figure 4d indicates that when using AQuaMet, the average number of handovers is lower than that of the baselines. As the baseline approaches are not designed using a tolerance level, the number of handovers remain fixed. However, AQuaMet makes handovers only when the throughput performance requirement on the serving link is not met within

the tolerance level. So, when this tolerance requirement is relaxed there is a slight decrease in the average number of handovers. For similar reasons as mentioned previously we have omitted the IATP baseline (average of 41.6 handovers) from this plot.

In summary, our initial evaluations demonstrate that a 116% improvement over SS and a 20% improvement over ATOM can be obtained for QoS violations when using the robust probabilistic model of AQuaMet. While, AQuaMet and IATP indicate similar QoS performance, IATP obtains this at the high cost of increase in number of handovers by 35 times. These demonstrated benefits are promising results for an extended evaluation in real environments.

VI. CONCLUSION AND FUTURE WORK

We have addressed the problem of representing the network state of dynamic multi-RAT networks using probabilistic abstractions in AQuaMet. This problem is at the forefront of 5G to realize the full potential of multiple coexisting RATs and connectivity options in densified networks. Previous contributions in this direction have been limited in their ability to address dynamic variations in user performance and drive management decisions with the objective of taming this variability. AQuaMet has been designed to overcome these challenges with a probabilistic abstraction of the network state based on downlink *attainable throughput* that quantifies performance variability in a RAT-agnostic manner, allowing controller applications to make intelligent decisions to satisfy guarantees on QoS. AQuaMet has also been designed for a distributed deployment making it scalable and computationally efficient. It uses existing metrics in existing standards making it practically applicable in current networks.

We verified the applicability of AQuaMet in controller decisions by coordinating downlink probabilistic interface selection, based on AQuaMet in a WiFi-LTE heterogeneous network. AQuaMet results in fewer performance violations compared to ATOM (20%), significantly fewer compared to SS (116%) and also reduces handover cost compared to IATP (by 35 times). This improvement in performance with a reduction in handover cost has a direct impact on the efficiency of resource utilization in the network as well as user satisfaction. We are currently working on extending AQuaMet with additional RAT-agnostic QoS estimators (e.g. delay). Future work includes extended applicability to other RATs (e.g. mmWave) and uplink modelling along with verification in test-bed networks.

ACKNOWLEDGEMENT

REFERENCES

[1] "5G technology etsi," http://www.etsi.org/technologies-clusters/technologies/5g, accessed: 2017-10-28.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5g be?" *IEEE Journal on selected areas in communications*, vol. 32, no. 6, 2014.

[3] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, 2016.

[4] A. Kostopoulos, G. Agapiou, F.-C. Kuo, K. Pentikousis, A. Cipriano, D. Panaitopol, D. Marandin, K. Kowalik, K. Alexandris, C.-Y. Chang et al., "Scenarios for 5g networks: The coherent approach," in *Telecommunications (ICT), 2016 23rd International Conference on*. IEEE, 2016.

[5] A. Orsino, G. Araniti, A. Molinaro, and A. Iera, "Effective rat selection approach for 5g dense wireless networks," in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*. IEEE, 2015.

[6] R. Mahindra, H. Viswanathan, K. Sundaresan, M. Y. Arslan, and S. Rangarajan, "A practical traffic management system for integrated lte-wifi networks," in *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 2014.

[7] U. P. Moravapalle, S. Sanadhya, A. Parate, and K.-H. Kim, "Pulsar: improving throughput estimation in enterprise lte small cells," in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. ACM, 2015.

[8] L. De Cicco, S. Mascolo, and V. Palmisano, "Skype video responsiveness to bandwidth variations," in *Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video*. ACM, 2008.

[9] S. Deb, K. Nagaraj, and V. Srinivasan, "Mota: Engineering an operator agnostic mobile service," in *Proceedings of the 17th annual international conference on Mobile computing and networking*. ACM, 2011.

[10] L. C. Gimenez, I. Z. Kovács, J. Wigard, and K. I. Pedersen, "Throughput-based traffic steering in lte-advanced hetnet deployments," in *Vehicular Technology Conference (VTC Fall), 2015 IEEE 82nd*. IEEE, 2015.

[11] S. Singh, S.-p. Yeh, N. Himayat, and S. Talwar, "Optimal traffic aggregation in multi-rat heterogeneous wireless networks," in *Communications Workshops (ICC), 2016 IEEE International Conference on*. IEEE, 2016.

[12] K. Ahuja, B. Singh, and R. Khanna, "Network selection based on available link bandwidth in multi-access networks," *Digital Communications and Networks*, vol. 2, no. 1, 2016.

[13] S. Deng, A. Sivaraman, and H. Balakrishnan, "All your network are belong to us: A transport framework for mobile network selection," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*. ACM, 2014.

[14] X. Yan, P. Dong, T. Zheng, and H. Zhang, "Fuzzy and utility based network selection for heterogeneous networks in high-speed railway," *Wireless Communications and Mobile Computing*, vol. 2017, 2017.

[15] A. Chakraborty, S. Sanadhya, S. R. Das, D. Kim, and K.-H. Kim, "Exbox: Experience management middlebox for wireless networks," in *Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies*. ACM, 2016.

[16] J. Herzen, H. Lundgren, and N. Hegde, "Learning wi-fi performance," in *Sensing, Communication, and Networking (SECON), 2015 12th Annual IEEE International Conference on*. IEEE, 2015.

[17] P. Torres Compta, F. H. Fitzek, and D. E. Lucani, "Network coding is the 5g key enabling technology: Effects and strategies to manage heterogeneous packet lengths," *Transactions on Emerging Telecommunications Technologies*, vol. 26, no. 1, 2015.

[18] "802.11k wifi standard," https://standards.ieee.org/findstds/standard/802.11k-2008.html, accessed: 2017-10-28.

[19] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive mac protocol for multi-hop wireless networks," in *Proceedings of the 7th annual international conference on Mobile computing and networking*. ACM, 2001.

[20] M. T. Kawser, N. I. B. Hamid, M. N. Hasan, M. S. Alam, and M. M. Rahman, "Downlink snr to cqi mapping for different multipleantenna techniques in lte," *International Journal of Information and Electronics Engineering*, vol. 2, no. 5, 2012.

[21] S. Sesia, M. Baker, and I. Toufik, *LTE-the UMTS long term evolution: from theory to practice*. John Wiley & Sons, 2011.

[22] "NS3 network simulator," https://www.nsnam.org/, accessed: 2017-10-28.

[23] L. Toni, R. Aparicio-Pardo, G. Simon, A. Blanc, and P. Frossard, "Optimal set of video representations in adaptive streaming," in *Proceedings of the 5th ACM Multimedia Systems Conference*. ACM, 2014.