

On design challenges of an endpoint flow association optimization service in a multi-provider wireless heterogeneous network

Jianwei Liu[†], Xin Xing^{*}, Kang Chen^{*}, and James Martin[†]

[†]School of Computing, Clemson University, Clemson, SC 29634

Email: {jianwel, jmarty}@clemson.edu

^{*}Department of Electrical and Computer Engineering, Southern Illinois University, Carbondale, IL 62901

Email: {xin.xing, kchen}@siu.edu

Abstract—Resource allocation optimization is critical to the overall performance of wireless heterogeneous networks (HetNet). Endpoint flow association optimization tries to reach a target resource allocation outcome by only controlling the associations of data flows to abstract/concrete network interfaces at the two ends of communication sessions. An instance of this problem in a single cellular heterogeneous network is the user association optimization problem, which plans the usage of macrocells and picocells. In this paper, we focus on this problem in a multi-provider HetNet, where the component wireless networks belong to different organizations. We identify and abstract the problem, comparing with the similar systems that are under the control of a single provider. After analyzing the changes in the new context and their implications, we provide a general centralized over-the-top network service design, which does not assume controls to the underlying network infrastructure. With the problem and design in mind, we then explore the performance problem when extending the existing user association schemes in a single-provider context to the new design. Though various association schemes are proposed in the prior research under different contexts, few of them provide an in-depth evaluation to several fundamental problems that are required by the problem in single/multi-provider context, i.e. 1) the distances of the association schemes to the optimal solution under various scenarios; 2) the sources and impacts of the potential throughput estimation errors. By using relatively small scale and more controlled scenarios, this paper is the first to provide answers to the above questions, which are valuable for guiding further studies and real system designs.

I. INTRODUCTION

Resource allocation (RA) optimization of wireless systems usually involves modeling the system in an abstract manner and formulates the problem to address the compromises between system and user oriented objectives such as spectral efficiency and throughput fairness. One way to implement this optimization is by controlling the user associations to Access Point (AP) or Base Stations (BS) at the client devices. A number of recent publications have proposed solutions for this type of optimization problem in a cellular wireless heterogeneous network (HetNet). For example, [1] has identified the problem in a 3G wireless network where macrocells and picocells co-exist and overlap. We first use Fig. 1 to illustrate this type of heterogeneous wireless network. The networks S_i in the figure are picocells, while M is a macrocell. They are operated by the same provider and share the same gateway and backhaul network (G) towards the outside Internet. We call it a

Single-Provider HetNet (SP-HetNet). With these assumptions, [1] identified the user association optimization problem which tries to achieve the *overall/generalized proportional fairness (PF)* in such a mixed network when all the BSs use PF schedulers. They assume backlogged downstream traffic to all the UEs. Fundamental to the modeling is the assumption that the PF is equivalent to maximizing a utility function of user throughputs when there is a single bottleneck in the network to optimize [2]. The problem is NP-hard and the authors propose offline optimal solutions and online greedy heuristics. The work in [3] provides a centralized solution based on convex optimization, and a distributed scheme to approximate it. In common, these SP-HetNet literature assumes this association controller resides in the radio network controller (RNC) of the backhaul network, and the control plane can be added to the messaging protocol between the BSs and the RNC. Though their preliminary results show improved spectral efficiency and fairness from a global perspective, the following problem remains untouched in these works.

Their methodology assumes the throughput model is accurate, which ignores the errors from a) inaccurate throughput model that cannot exactly reflect the real path loss, fading and resource division, etc.; b) difference between system dynamics and control frequencies; c) impact from the backhaul network. This paper tries to identify these problems for the first time and shows the impacts of them to the existing algorithms.

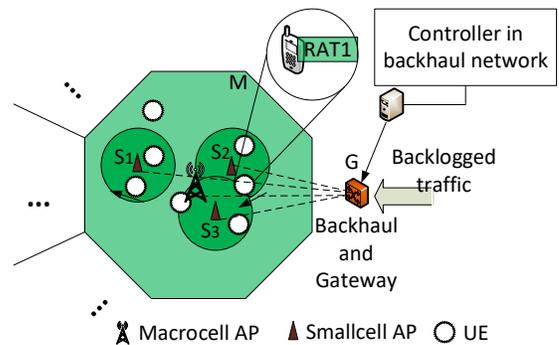


Fig. 1. System contexts for a Single-RAT SP-HetNet.

Meanwhile, nowadays one location is usually covered with multiple wireless access networks that belong to multiple

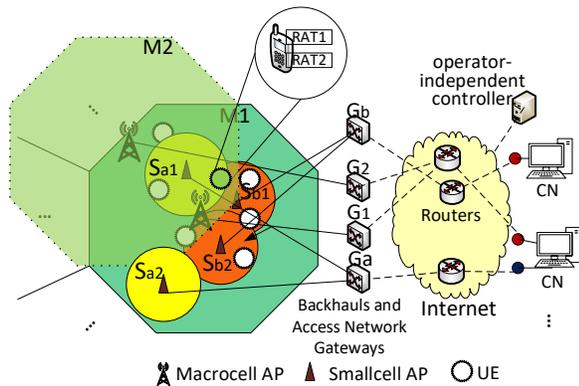


Fig. 2. System contexts for MP-HetNet.

providers. The newly developed network stacks and applications ([4], [5]) can enable each mobile application to use multiple wireless interfaces concurrently. All the pieces exist to build large scale multi-provider heterogeneous networks (MP-HetNet) that support multi-homed mobile devices. Recent literature [6], [7] considers such systems. Fig. 2 illustrates a hypothetical MP-HetNet. We depict four *component wireless networks* (CWN) in the figure. The networks M_i represents two cellular networks by different operators; the networks S_{ai} and S_{bi} are two groups of WiFi networks that use the same Internet Service Provider. Each UE owns multiple wireless interfaces. Since CWNs are overlapped, each interface of UE can connect to one of the CWNs with a matching Radio Access Technology (RAT) at a time. Therefore, a user association scheduling problem similar to that in the SP-HetNet exists. In addition, this new HetNet owns the following features, 1) Every CWN owns a unique gateway and backhaul network. For example, a Sprint LTE network and a campus WiFi will certainly have separated gateways and backhaul networks; 2) The individual APs (we use AP to generalize AP and BS in this paper) can use heterogeneous schedulers, or even sometimes unknown schedulers. For example, most of the previous literature in single-RAT HetNet assumes PF schedulers. However, WiFi networks usually use throughput fairness schedulers. Because of these differences, we need to relook at the flow association optimization problem. We can at least think of the following four changes, 1) by aggregating multiple wireless interfaces, it is more likely device will experience congestion at the backhaul network; 2) the heterogeneity of backhaul network make it impossible to assume homogeneous backhaul performance like the RTT and loss rate, which are essential for end-to-end throughputs; 3) the controller in design should be operator independent, i.e. it cannot reside in any operator's backhaul network and should be accessible by any UE. Further design of a flow association optimization framework in a MP-HetNet requires careful examination to these changes.

Therefore, we think of an Over-The-Top (OTT) style flow association engineering frameworks. (We use the term *flow association* to indicate the control granularity can be finer than user association, e.g. in application or socket connection level.) By OTT, we mean the design is against the previous resource allocation frameworks which assume controls/modifications

to the network components, such as [8], [9], [10]. It only assumes control to the ends of communication sessions. The previous user association based literature in SP-HetNet can be considered as a special case of this general OTT style flow association engineering framework. This paper generalizes and abstracts the problem, and study some critical design issues that are common in both SP-HetNets and MP-HetNets when designing an OTT flow association engineering framework.

This paper focuses on the following questions that are valid in both HetNet environments, but that are likely more problematic in a MP-HetNet. 1) Even assuming an accurate throughput estimation model with no error, what are the distances of the previous association schemes to the optimal and status quo policy based association schemes, in terms of the PF objective function value and other global performance metrics? 2) Whether the input errors exist, and what are their impacts to the performance of the solutions? Similar to the previous literature [1], we focus on the optimization problem of **downstream** like video streaming from CNs to UEs, as this type of traffic takes up most of the Internet traffic. Upstream is much more difficult when channel condition is used for throughput estimation as in [1], [3]. However, the measurement based throughput estimation method in [11] can help to alleviate this problem under our system design. Another crucial aspect of the modeling is the assumption of the percentage of participating devices. In this paper, we assume full deployment where the system can control all the UEs under every AP.

The contributions of this paper include, 1) identifying for the first time the OTT resource allocation/traffic engineering problem in a MP-HetNet, and explaining why an OTT design is required; 2) a new methodology for an in-depth examination of the performance distance of previous schemes to the status quo scheme and the optimal solution; 3) testing various methods' sensitivities to the input errors and system parameter selection. The results can serve as guidance for the design and implementation of a real OTT resource allocation framework.

This paper is organized as follows. We review the previous literature in section II. Then we present an abstract mathematical model of the flow association schemes, and a generalized system architecture for both OTT and non-OTT solutions in section III. In section IV, we design new methodology to examine the performance distance of the previous schemes to the policy-based scheme used in real system, and the optimal solution. In section V, we explore the causes of the input error to scheduling algorithms and prove its existence by simulation. We then test the sensitivities of various schemes to general errors and the error from control frequency. Finally, we make our conclusions in section VI.

II. RELATED WORKS

Though a lot of previous literature on HetNet resource allocation problem [1], [3], [12], [6], [7], [9], [8] with some of it OTT type [1], [3], [6], [7], there is no sufficient modeling work that researches the fundamental problems in such an OTT type flow association service.

In general, these works use the generalized proportional fairness as the overall optimization objective, and focus on backlogged downlink traffic. Most of them assume the individual schedulers use proportional fairness. Since the problem is proved to be NP-hard [6], [1], greedy heuristic and approximation dominate the prior research. However, they fail to provide a detailed comparison to the current policy based scheme used on smart phones, and the optimal solution. Also, there is no existing report of the comparison among the previous schemes. This paper tries to provide a more in-depth view of the problem itself, the comparison of different methods, and the impact of various system dynamics and parameter selections.

III. ABSTRACT MODEL AND SOLUTION ARCHITECTURE OF THE FLOW ASSOCIATION PROBLEM

The study in this paper is based on the following mathematical model, which is similar to the ones in [1], [3], [7], [6]. The models in the previous literature in general assumes, 1) generalized proportional fairness as the overall optimization goal, as PF is considered a simple and fair trade-off between the aggregated throughput and fairness [1], [13], [14]; 2) backlogged traffic or infinite application layer demand; 3) rough association control granularity.

$$\begin{aligned} & \text{Maximize} && \sum_{j=1 \dots M} \sum_{i=1 \dots N} \log(T_{ij}) * x_{ij} \\ & \text{subject to} && \sum_j x_{ij} = 1, \\ & && T_{ij} = \mathcal{U}(\hat{x}, \dots), \\ & && x_{ij} \in \{0, 1\} \end{aligned} \quad (1)$$

The index i is for UE, while j is for AP. N is the total number of UEs, while M is that for APs. x_{ij} is the association variable that decides whether the user i should connect to AP j . It is either 0 or 1, which is called *integral solution*. The solution to the above problem is a matrix of x_{ij} (\hat{x}), which is the flow association guideline provided to UEs by the framework. The first constraint means every UE can only connect to exactly one interface. T_{ij} is the end-to-end throughput of UE i if it is connected to AP j . The second constraint shows that the end-to-end throughput is a function that involves the association plan \hat{x} and other factors. Those can include the signal-to-noise-ratio of a link and the other end-to-end connection characteristics, like loss rate and delay. The function \mathcal{U} abstracts the impacts of both the scheduling scheme of individual APs and the end-to-end dynamics. This model is similar to the model in Eq. (1) - (4) in [1], but with a more generalized form.

To solve the problem with an either OTT or non-OTT style, we assume a flow association engineering framework should at least have four core components, i.e. 1) A regional resource controller (RRC); 2) A flow scheduling information collection module; 3) A control plain that delivers the messaging such as scheduling information and policies; 4) A scheduling policy enforcement module. Fig. 3 shows these components and their interactions. Note, in a non-OTT style solution, previous

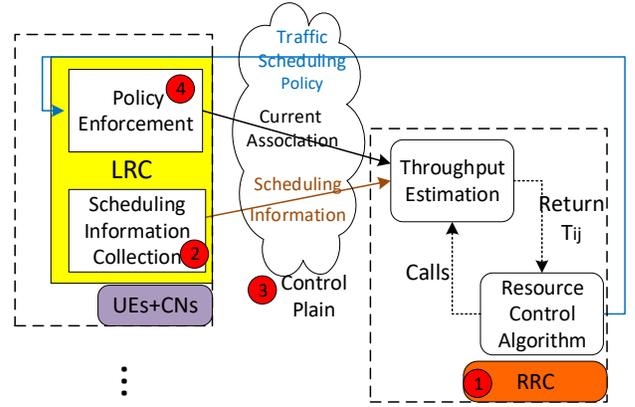


Fig. 3. OTT architecture of a flow association engineering framework.

literature assumes the RRC resides in the backhaul network, and the control plain relies on the extensions of the 3GPP control messaging protocols. The components 2) and 4) could be implemented in any network node. In contrast, in an OTT solution, the RRC has to be operator-independent, and the components 2) and 4) can only work on the two ends of communication session, i.e. UEs and CNs. These changes also mandate a control plain redesign. We call the components 2) and 4) at the end devices an Local Resource Controller (LRC). As shown in Fig. 3, the LRCs collect the current association and connection status information from UEs and sends it to the RRC. The CNs may optimally send the measured connection status information to the connected UE. The RRC then runs the scheduling algorithm which calls the throughput estimation module. The throughput estimation module estimates the throughput on every interface of UE under the planned association policy from the scheduling algorithm. The output of the scheduling algorithm is an association policy which optimizes certain system-wide performance objective. Every LRC receives its scheduling plan periodically from the RRC and enforces it locally. Though we present the RRC as a centralized service, its storage and computation can be distributed. It is only conceptually a centralized service. More feasibility analysis of this design can be found in [11].

IV. BASELINE STUDY: PERFORMANCE OF THE EXISTING USER ASSOCIATION SCHEMES

Given the above optimization problem in Eq. 1 and the system design in section III, the first question is, how well will the scheduling algorithms in the previous work behave when compared with the status quo association scheme and the optimal solution. Specifically, we are interested in the scenarios where all the scheduling algorithms have a perfect input, i.e. 100% accurate throughput estimation (which is generally assumed in the previous literature). In this section, we try to provide these information using a representative set of previous methods and a new methodology.

First, we set up a scenario that has N UEs and M APs. We test with $N = \{5, 10\}$ and $M = 3$. The scales of M and N are purposely kept small so that the optimal solution is tractable. To make the percentage of UEs with good connection to each

AP more controllable, we do not use random mobility as in previous literature [1], [6]. Instead, we control the percentage of users connecting to each AP. One AP with large coverage (like a LTE macrocell) and several APs with smaller coverage (like a WiFi AP) coexist. We control the percentage of UEs under the small cells (P_s). The P_s values we tested are {0.8, 0.6, 0.4, 0.2}. For each time slot, we generate the distance to APs randomly following the P_s .

Also, we test with two cases concerning the type of APs:

Case 1. All proportional fair APs. (as in [1], [3], shorten for 'pf only' in the figures).

Case 2. A mix of proportional fair and throughput fair APs. (as in [7], [12], shorten for 'both' in the figures).

This assumption actually has significant implication to the modeling. With the former, the T_{ij} in Eq. 1 is concave and the problem is convex after applying the logarithm function. Therefore, we can use convex solvers (e.g. the interior point method solver in MATLAB) to solve the problem easily. However, in Case 2, with the added complexity of throughput fairness scheduling APs, the T_{ij} is not concave anymore, rendering the convex solvers unusable. We will show its impact to the convex solver based methods (the first centralized method in [3]) in the results below. In the simulation, we use the distance to throughput mapping we measured from NS3 [15] for the peak rates before resource divisions. The detailed mapping is provided in the appendix of the full version paper [16].

We compare the following five methods.

- 1) policy-based: This represents the interface selection scheme on most of smart phones, i.e. when WiFi is available, connect to the WiFi with the best signal; otherwise, connect to the cellular network.
- 2) optimal-brute-force: An optimal solution using a brute-force method which iterates through all the possible user association configurations, and returns the one with the largest objective function value.
- 3) gpfHeuristic: Greedy algorithm generalized from the Algorithm 1 in [7], which is shown in the Algorithm 1 below. We generalize the algorithm in [7] to m LTE BSs instead of a single BS and select k WiFi APs to offload instead of one. The basic idea is that every macrocell forms a set containing UEs that can connect to it. For each set s , we first initialize all the UEs in it to the macrocell, and then try to offload them to k WiFi APs. From the n^k cases, we select the one that results in the maximum objective function value, and finalize the UEs to the corresponding small cells. The algorithm repeats this process until all the sets are checked.
- 4) random: Returning the best solution in 5 randomly generated user association configurations.
- 5) round-off-interior-point: Solving the problem using a non-linear solver (like *fmincon* in MATLAB), and return the round-off integral solution.

We compare the following three metrics,

- 1) PF value: The objective function value in Eq. 1.

Algorithm 1: Generalized greedy heuristic.

```

1 for  $u$  in  $\mathbb{U}$  do
2   if  $R_{ij} > 0$  then
3      $u$  into set  $S_m$ ;
4   end
5 end
6 for for  $s \in S$  do
7   select  $k$  small cells and try to offload UEs in  $s$  to the small cells ( $n^k$ 
   scenarios in total);
8   select the scenario above that results in the maximum object value, and
   finalize those UEs to the small cells;
9 end

```

2) Spectral efficiency: Sum of all the user throughputs divided by the overall spectrum used.

3) Jain's fairness metric: $J(T_1, T_2, \dots, T_n) = \frac{(\sum_{i=1 \dots n} \frac{T_i}{T_{b_i}})^2}{n * \sum_{i=1 \dots n} (\frac{T_i}{T_{b_i}})^2}$, where T_i is the throughput of user i , and T_{b_i} is the throughput of user i of the optimal global proportional fairness solution (e.g. the one generated by the brute-force method below).

Fig. 4 shows the results for Case 1, i.e. only PF scheduling APs, while Fig. 5 is the results for Case 2, i.e. both PF scheduling APs and throughput fairness APs.

We have the following observations from these figures,

- (1) Concerning PF value, optimal-brute-force always produces the best result for both cases, which serves as the baseline and a sanity check. In Case 1, round-off-int has close-to-optimal performance most of time. However, in some rare cases the round-off-int can produce remarkably worse configuration that has only half of the optimal PF value, which is even worse than the policy and random based methods. This is perhaps a result of local optimality.
- (2) In both cases, gpfHeuristic is the second-closest to the optimal in terms of PF value and spectral efficiency, and with lower standard deviation compared with random and policy based. This means, though not optimal, the heuristics like gpfHeuristic work pretty well. However, it scarifies the fairness and has much larger standard deviations compared with the optimal in both throughput and fairness metrics.
- (3) The methods have similar Jain's fairness index performance under the Case 2. But in Case 1, the layers among different methods are relatively clear. We can see basically optimal < round-off-int \approx policy < gpfHeuristic < random. All methods except the optimal have large standard deviation in terms of Jain's fairness index.
- (4) The random method has the worst performance and the largest deviation most of time. The policy based method is only slightly better than the random generated configurations.

The result in this section shows, even with perfect information, the current algorithms will have some distance to the optimal global PF objective. This has very important implication for the study below about the impact of the input error to the algorithms, i.e. when estimating the impact of input error in the section V, we need to start with the optimal-brute-force solution. Because the results of the optimal solution alone have no influence from the error introduced by algorithms.

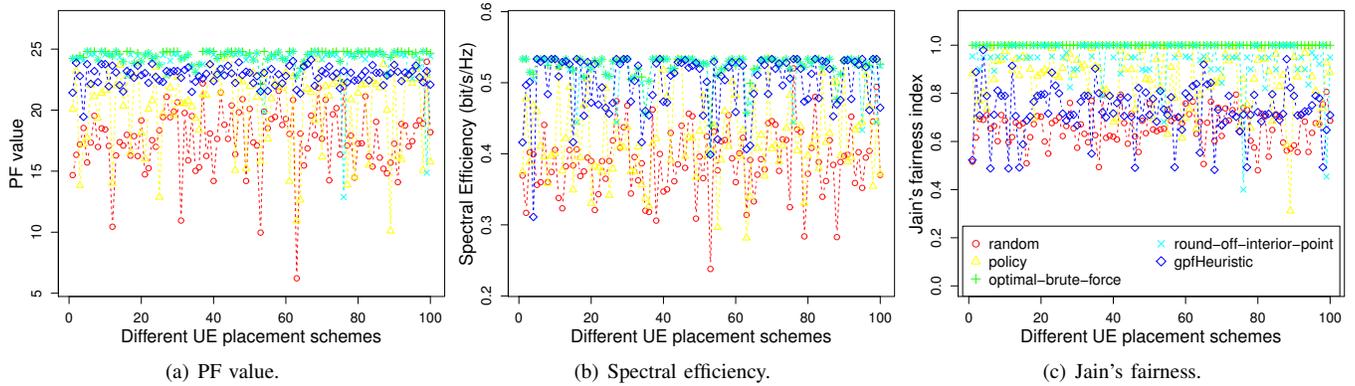


Fig. 4. The distance of various previous schemes to the optimal (10 UEs and 3 APs, $P_s = 0.8$, pf only).

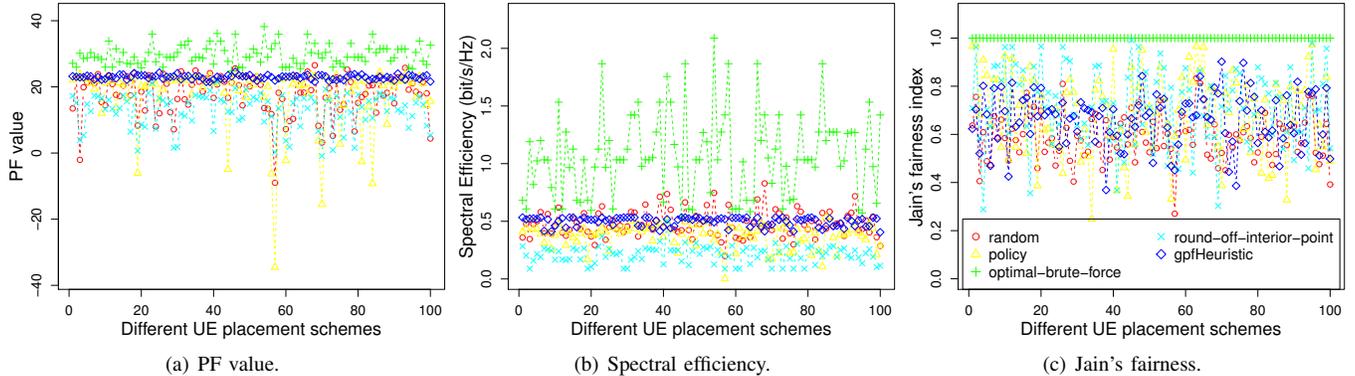


Fig. 5. The distance of various previous schemes to the optimal (10 UEs and 3 APs, $P_s = 0.8$, both).

V. SOURCES AND IMPACTS OF INPUT ERRORS

In the last section, we use a similar assumption as those in the previous literature, i.e. there is an explicit formula that can estimate user throughput accurately. For example, in [3], they use the Shannon equation. However, this assumption is invalid in real systems. In general, the throughput estimation errors can come from,

- 1) **Type I:** throughput model, which includes,
 - a) capacity and individual scheduler modeling error;
 - b) failing to consider application demands in the model;
 - c) failing to include impacts from the backhaul networks.
- 2) **Type II:** the change of inputs (e.g. connection status and association information) between sampling when the plan is enforced at the LRC. It consists of two phases, i.e. 1) from sampling to RRC scheduling 2) from RRC scheduling to local policy enforcement.

Meanwhile, as shown in Fig. 3, the throughput estimation model provides input to the resource control (a generalization of the controls to association and rate) algorithm. However, the rough throughput estimation models used in the previous literature inevitably introduce errors to this input. None of the prior work studies the sensitivity of their algorithms to the input errors, i.e., assuming the throughput model has a 100% accuracy in simulations. Therefore, in this section, we study the sensitivities of some representative user association schemes to various levels of input errors. We would select the same set of algorithms as in section IV. However, the problem is invalid for the policy based and the random methods since they do not use throughput estimation. Thus, we only compare

three schemes in this section, i.e. 1) optimal-brute-force; 2) gpfHeuristic; 3) round-off-interior-point. The subsection V-A deals with the Type I error, while the subsection V-A2 builds simplified model to test the impact of the Type II errors.

A. Type I error

1) *Preliminary experiments to show the errors exist:* Due to the limitation of space, we only show why the rough throughput estimation model based on Shannon capacity equation (like in [3]) will introduce Type I error. We use the following simulation in NS3 to demonstrate the concept. During the simulation, we position the UE at various distances to an LTE Base Station. The UE has one dominating downstream application. We use the following methods to estimate the throughputs of the UE and compare with the TCP and UDP throughputs measured by *iperf*. It is tested with two scenarios, i.e. 1) no fading; 2) with a more realistic fading model based on the Annex B.2 of 3GPP TS36.104 [17]. We use pedestrian model with a speed of 3kmph.

- 1) Shannon Equation: $T = B * \log_2(1 + SNR)$, where B is the bandwidth of the spectrum in use, and T is the estimated user throughput. SNR represents the downlink signal to noise ratio, i.e. $SNR = \frac{SP}{NP}$, where SP is the power of signal and NP is the power of noise (in watt).
- 2) Modulation and Coding Scheme (MCS). For example, if 64QAM is used, every symbol has 6bits. If c symbols can be supported for a 10MHz channel, the prediction result will be $6c$ bps.
- 3) Nyquist Equation: $T = 2B * \log_2(Nbits)$, Where $Nbits$ is the number of bits used for the coding scheme.

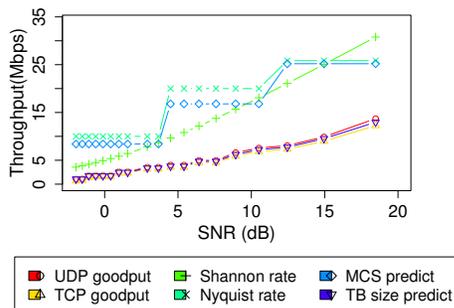


Fig. 6. Predicted and the measured throughputs

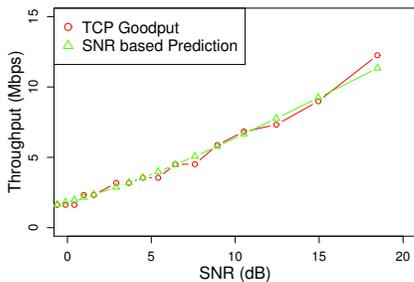


Fig. 7. Scaled predicted throughput vs the measured throughput (no fading).

- 4) Transport Block (TB) size. A MAC layer rate limiter in LTE, which sets an upper bound for the maximum bytes one UE can transmit given a specific MCS and number of resource blocks. The TB size is based on the table in 3GPP standard TS36.101, Annex A.2.1.2 [18].

Fig. 6 shows the distance of the raw results using the above throughput estimation methods to the goodputs. Fig. 7 shows the best curve fitting of SNR based estimation curves to the TCP goodput when no fading added; while Fig. 8 shows that when the fading model added.

From the results, we can see that the throughput estimation, even with the best fitting parameters, can deviate from the measured goodput, especially with the case of realistic fading. More importantly, the directions of this deviation are random. Therefore, in real system, the errors cannot be easily offset by a constant factor. This at least proves that the Type I throughput estimation error exists using the previous SNR based throughput estimation methods.

2) *Conceptual error sensitivity tests:* In this section, we test the sensitivities of the previous user flow association schemes to the input errors. We set up a similar scenario like in section IV. However, this time we purposely insert errors to the inputs of the algorithms, i.e. the estimated user throughput (T). The new throughput after error insertion $T' = (1 \pm e) * T$, where e is the error rate. We use $e = \{0.05, 0.1, 0.2, 0.3, 0.4,$

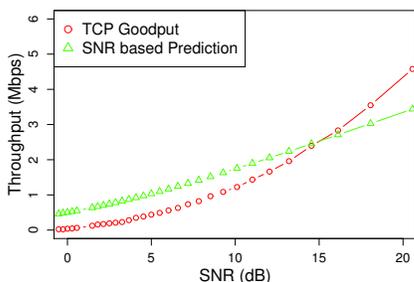


Fig. 8. Scaled predicted throughput vs the measured throughput (with fading).

0.5} in our evaluation. Every experiment lasts for 10000 time slots. Fig. 9 shows the results for the method optimal-brute-force. We first show the impact to this method, because only its results do not contain the errors from the method itself. We provide the results of the other methods in the full version of the paper [16]. From Fig. 9, we observe that,

- (1) For the optimal solution, the input error will always lead to degraded PF objective function value (shown in Fig. 9(a)). Meanwhile, the percentage of performance degradation is nearly proportional to the error rate inserted.
- (2) The system performance in terms of both aggregated throughput and fairness will have larger standard deviation after errors are inserted. The performance can get better or worse with much larger percentage compared with the PF value. This is because the proportional fairness is only a tradeoff between the two optimization objectives. There is still space that either or both metrics can be improved [14]. This is the artifact of the proportional fairness itself. However, as a widely used and simple resource allocation optimization objective, this paper is more interested in how it behaves, instead of whether a better objective can be used.
- (3) From Fig. 9(b), we observe that the aggregated throughput may both increase and decrease; and its CDF curve is almost symmetric. This means, in all the error levels (up to 50% error), the optimal-brute-force has almost equal chance to achieve a degraded throughput, as it can get a better one. Different from that, from Fig. 9(c), the Jain's fairness index's CDF is biased to the right side of the x axis, which means the method has larger chance to get a better solution in terms of Jain's fairness index. This is reasonable because proportional fairness gives aggregated throughput more weight. When input errors are inserted, some random association configurations are given. Under this scenario, the aggregated throughput has less room to be improved. Because it is closer to the optimal compared with the fairness index.

B. Type II errors

We simulate and verify the impact of the Type II errors in the following way. Assuming an atomic time unit (t_a) of SNR change, we vary the control time interval (t_c) as multiples of this time unit. The multiples α we tried are {1, 3, 7, 15, 31}. We continue to use $N=5$, $M=3$, and the WiFi coverage rate $P_s=0.8$. For the results with more P_s values, please refer to the appendix in the full version of the paper [16]. We test with SNR change rate (C_r) = {0.1, 0.3, 0.5}. $C_r = \frac{SNR_{t_t}}{SNR_{t-t-1}}$ where t represents the current time slot. Figs. 10 shows the results for two methods under the Case 1 APs. From the results, we have the following observations,

- (1) When the control time interval is larger than 10 times of t_a ($\alpha = \{15, 31\}$), the optimization objective value starts to degrade. For the scenarios with a smaller t_c , the performance degradation is negligible. This implies, in real system design and implementation, we most likely want to use control frequency that is less than 10 times of average SNR change frequency, and be cautious about the opposite case.

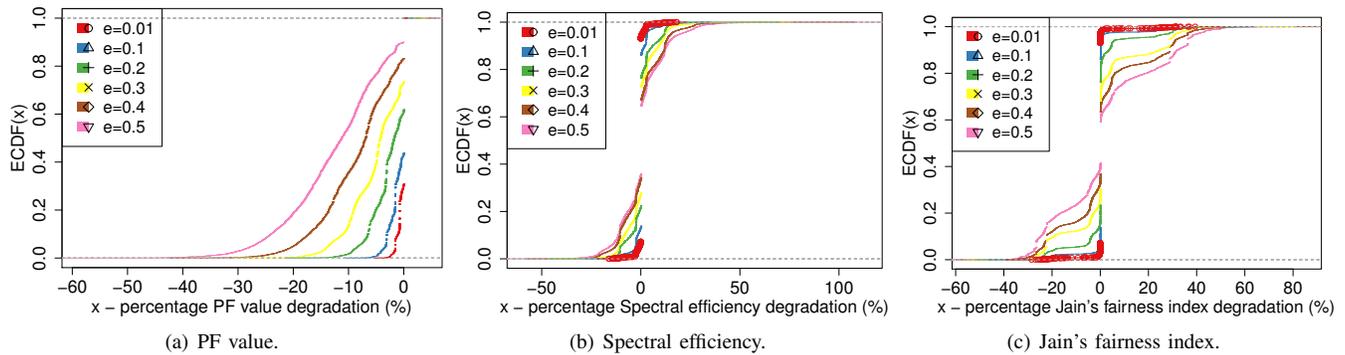


Fig. 9. Impact of throughput estimation error to optimal-brute-force (5 UEs and 3 APs, $P_s = 0.8$, pf only).

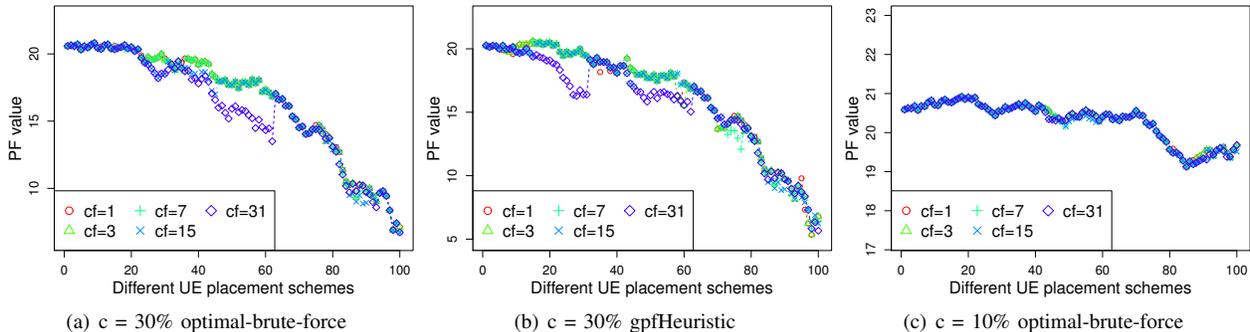


Fig. 10. Impact of control frequency to the PF value of various methods (5 UEs and 3 APs, $P_s = 0.8$, pf only).

(2) The impact of control frequency is also related to the SNR change rate. For example, for SNR change rate of 10%, larger control interval will have little impact.

VI. CONCLUSION

From the analysis and simulation results, we can see the input throughput error exists, which is more problematic in a MP-HetNet. We also observe that directly applying the existing user association schemes in SP-HetNets to the new context can result in performance degradation and large deviations. Additionally, we tested the sensitivities of the existing schemes to the errors introduced by system control frequency. From the results, we learn the range of usable control frequencies without drastic performance degradation. These insights can serve as the guidance for future system design and implementation for a real OTT optimization framework. For example, knowing the existence and impacts of input errors, we can use measurement based method to reduce the throughput estimation error as we proposed in [11]. In the future, we would like to investigate the scenarios when multiple interfaces can be simultaneously utilized by UEs, and the scenarios with various control granularities.

REFERENCES

- [1] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *Proc. of INFOCOM*, April 2006, pp. 1–12.
- [2] F. P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, 1997.
- [3] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. Andrews, "User association for load balancing in heterogeneous cellular networks," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 6, pp. 2706–2716, June 2013.
- [4] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure, "Rfc6824: Tcp extensions for multipath operation with multiple addresses," 2013.
- [5] A. L. Ramaboli, O. E. Falowo, and A. H. Chan, "Bandwidth aggregation in heterogeneous wireless networks: A survey of current approaches and issues," *Journal of Network and Computer Applications*, vol. 35, no. 6, pp. 1674–1690, 2012.
- [6] S. Deb, K. Nagaraj, and V. Srinivasan, "Mota: Engineering an operator agnostic mobile service," in *Proc. of MobiCom*. New York, NY, USA: ACM, 2011, pp. 133–144.
- [7] R. Mahindra, H. Viswanathan, K. Sundaresan, M. Y. Arslan, and S. Rangarajan, "A practical traffic management system for integrated lte-wifi networks," in *Proc. of MobiCom*. ACM, 2014, pp. 189–200.
- [8] R. Amin, J. Martin, J. Deaton, L. DaSilva, A. Hussien, and A. Eltawil, "Balancing spectral efficiency, energy consumption, and fairness in future heterogeneous wireless systems with reconfigurable devices," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 5, pp. 969–980, May 2013.
- [9] H. Zhang, F. Bai, and X. Ju, "Heterogeneous vehicular wireless networking: A theoretical perspective," in *Proc. of WCNC*. IEEE, 2015, pp. 1936–1941.
- [10] A. Anand and G. de Veciana, "Invited paper: Context-aware schedulers: Realizing quality of service/experience trade-offs for heterogeneous traffic mixes," *WiOpt*, 2016.
- [11] J. Liu, A. Rayamajhi, and J. Martin, "Using mptcp subflow association control for heterogeneous wireless network optimization," in *Proc. of WiOpt*, May 2016, pp. 1–8.
- [12] W. Wang, X. Liu, J. Vicente, and P. Mohapatra, "Integration gain of heterogeneous wifi/wimax networks," *Mobile Computing, IEEE Transactions on*, vol. 10, no. 8, pp. 1131–1143, Aug 2011.
- [13] D. Bertsimas, V. F. Farias, and N. Trichakis, "The price of fairness," *Operations research*, vol. 59, no. 1, pp. 17–31, 2011.
- [14] A. Sediq, R. Gohary, R. Schoenen, and H. Yanikomeroglu, "Optimal tradeoff between sum-rate efficiency and jain's fairness index in resource allocation," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 7, pp. 3496–3509, July 2013.
- [15] "NS3," <https://www.nsnam.org/>.
- [16] "Technical report on conceptual study of over-the-top resource allocation in heterogeneous wireless networks," <https://jianwei-liu.github.io/icc-full.pdf>.
- [17] "ETSI", "Base station (bs) radio transmission and reception, 3gpp ts 36.104 release 14," 2017.
- [18] ETSI, "User equipment (ue) radio transmission and reception, 3gpp ts 36.101 release 14," 2017.