On the Potentials of Traffic Steering in HetNet Deployments with Carrier Aggregation

Panagiotis Fotiadis, Michele Polignano
Aalborg University, Denmark
Department of Electronic Systems
E-mail: {paf, mpo}@es.aau.dk

Ingo Viering
Nomor Research GmbH
Munich, Germany
Email: viering@nomor.de

Paolo Zanier
Nokia Solutions & Networks
Munich, Germany
Email: paolo.zanier@nsn.com

Abstract—This paper investigates the potentials of Traffic Steering (TS) in Carrier Aggregation (CA) deployments. TS is defined as the mechanism to optimally distribute traffic among the deployed cells. Obviously, such functionality is rather crucial in scenarios where solely non-CA User Equipments (UE) are present. The introduction of CA is expected to simplify inter-layer load balancing, as CA UEs can concurrently connect to multiple carriers. In order to evaluate the relevance of TS in CA environments, a simple TS framework is developed, where load balancing decisions are applied in both Radio Resource Control (RRC) UE states. System level simulations have been conducted for different deployments and CA UE ratios. The corresponding results have shown that TS should be mandatory for any CA penetration below 50%, irrespective to the network deployment. Nevertheless, for scenarios where the number of deployed carriers is greater than the multi-carrier connectivity capabilities of the CA device, TS should be applied even for higher CA UE ratios, since the CA scheduler cannot solely resolve the load imbalances.

Index Terms—Traffic Steering; Radio Resource Management (RRM); Carrier Aggregation; Mobility Management

I. INTRODUCTION

The exponential growth of mobile broadband calls for adequate system upgrades in order to meet the future capacity targets [1]. As the spectral efficiency of Long Term Evolution (LTE) approaches the theoretical bounds, the sole densification of the macrocell layer will not be sufficient[2][3]. In fact, the deployment of low-power small cells (pico/ femtocells) along with additional LTE enhancements introduced in the LTE-Advanced standardization can further improve the spectral efficiency. Nevertheless, this complex cellular architecture raises new challenges in terms of mobility, interference and traffic management such to efficiently utilize system resources. Traffic Steering (TS) is denoted as the functionality to optimally distribute the load among the available cells subject to the aforementioned mobility and interference limitations.

In the context of small cell deployment and intra frequency TS, range extension techniques are often applied by virtually expanding the coverage footprint of low-power nodes [4]. In such a manner, enhanced load balancing is achieved and small cells are better utilized. However, the strong interference experienced by the small cell layer demands for Inter-Cell Interference Coordination (ICIC) [5] schemes in order to maximize offloading towards the low-power nodes. For inter-frequency TS, forced Handovers (HO) in the Radio Resource Control (RRC) Connected and specific reselection policies [6] in the RRC Idle can also be applied for boosting the resource consumption of an underutilized cell.

Carrier Aggregation (CA) [7] is one of the key LTE-Advanced features for further increasing trunking efficiency. CA User Equipments (UE) can concurrently connect to more than one carrier, also referred to as Component Carrier (CC) and, therefore, fast inter-layer load balancing can be achieved. More specifically, intelligent joint scheduling across the available CCs can actually balance the load between the different CCs, while treating both CA and legacy non-CA UEs in a fair manner [8]. In other words, load balancing can be provided solely by the scheduler functionality as long as a sufficient number of CA UEs exists in the network for exploiting CA.

In this paper, we focus on the interaction between TS and CA in Heterogeneous Network (HetNet) deployments. For that purpose, a load-based TS framework is developed and tested for different ratios of CA devices. Both RRC Idle and Connected UE states are assumed, while CA UEs support up to 2 serving CCs. As shown by Fig. 1, the study is conducted in 2 different Release 10 LTE-Advanced scenarios, implying that intra-site CA is enabled at the macrocell layers. The reason for considering both cases A and B is for evaluating the relevance of TS in CA environments, where the number of deployed CCs is equal or greater than the number of CCs that a CA UE supports.

The remainder of the paper is structured as follows. Section II provides a brief overview of the CA framework, while
the developed TS framework is described in Section III. The simulation assumptions along with the corresponding results are available in Section IV and V respectively. Finally, section VI concludes the paper.

II. CARRIER AGGREGATION FRAMEWORK

In CA mode, CA devices aggregate spectrum from more than a single CC. Backward compatibility with legacy non-CA users is guaranteed, since each LTE-Advanced CC follows the principle of conventional Release 8 LTE specifications. Therefore, link adaptation, Hybrid Automatic-Repeat-Request (HARQ) and transport block transmissions occur independently per CC. On the other hand, Radio Resource Management (RRM) functionalities such as Physical Resource Block (PRB) assignment and packet scheduling can be performed either independently per CC or jointly among different CCs. In particular, the joint co-operative approach enhances significantly the load balancing performance of CA, in scenarios where CA UEs can support concurrent transmissions from all deployed CCs. Such a scheduling scheme is also denoted as cross-CC Proportional Fair (PF) scheduler [8].

Regarding the set of serving cells, a CC is referred to as the Primary Cell (PCell) and is treated as such by the higher layer procedures [7]. Thus, amongst others, functionalities such as mobility management and RRC connection maintenance are solely performed at the PCell. PCell change is only possible via a HO execution. Since CA is only applicable at the RRC Connected state, a CA UE is assigned as a PCell the latest camping cell, whenever switching to RRC Connected. The remaining serving cells are denoted as Secondary Cells (SCells) and are dynamically added, removed or changed. Similarly to the non-CA UE mobility management framework, SCell actions are event-triggered, based on UE RRM measurements. Whenever a mobility event [9] is met, it is reported to the network and the base station performs the associated SCell action via RRC signalling. Nevertheless, depending on the UE multi CC connectivity capabilities, the availability of inter-frequency measurements differs. In principle, non-CA UEs initiate inter-frequency measurements only if the serving measurements drops below a pre-determined threshold (A2 event [9]). On the other hand, CA terminals periodically perform inter-frequency measurements without monitoring whether the A2 event condition is met. This measurement concept is referred to as Background Inter-Frequency Measurements (BIM) [10]. Nevertheless, the A2 event is still applicable for enabling inter-frequency HOs, regardless of the UE connectivity capabilities.

III. TRAFFIC STEERING SCHEMES

In this section, 2 different load-based TS schemes are presented for the RRC Connected and RRC Idle respectively. Inter-frequency TS driven events are triggered solely by the serving cell load conditions. Given a target bit rate $R_{target}$, the load contribution of a single user $u$ to cell $m$ derives as follows:

$$\rho_u = \min\left\{ \frac{f_u \cdot R_{target}}{R_u \cdot B_m} , \rho_{max} \right\}, \quad (1)$$

where $R_u$ is the actual bit rate that user $u$ achieves for being assigned $f_u$ resources in cell $m$. The cell bandwidth is denoted as $B_m$, while $\rho_{max}$ represents the maximum allowable load that a user can contribute to the cell. $\rho_{max}$ avoids situations where a single UE in bad Signal-to-Interference plus Noise (SINR) conditions can declare a cell as overloaded. For CA UEs, $R_u$ is the aggregated user-perceived throughput that user $u$ experiences in all CCs. Obviously, the overall load in cell $m$ is expressed as:

$$\rho_m = \min\left\{ \sum_{u=1}^{n} \rho_u , 1 \right\}, \quad (2)$$

The load conditions of neighboring cells are signaled in terms of Composite Available Capacity (CAC). CAC is denoted as the amount of resources that a cell has available for load balancing [11]. Thus, given a target operational load, $\rho_{target}$, the CAC of cell $m$ is modeled as:

$$CAC_m = \frac{B_m}{B_{max}} \cdot (1 - \frac{\tilde{\rho}_m}{\rho_{target}}), \quad (3)$$

where $\tilde{\rho}_m$ is the filtered cell load and $B_{max}$ is the maximum bandwidth allocation in the set of neighboring cells. The bandwidth ratio, $B_m/B_{max}$, represents the Cell Capacity Class Value (CCCV) [11].

A. TS upon Connection Establishment

UEs switching to the RRC Connected are explicitly requested to trigger inter-frequency measurements in terms of Reference Signal Received Power (RSRP) and Signal Quality (RSRQ), if the serving load exceeds the overload detection threshold $\rho_{high} > \rho_{target}$. For each CC, the strongest RSRP cell is reported subject to both minimum coverage and interference requirements. Among the set of load balancing candidate cells $\{p\}$, a forced HO is initiated towards the one with the highest CAC value. $RSRP_{min}$ is set $\Delta$ dB higher than the associated A2 threshold. In such a manner, inter-frequency ping pong HOs can be avoided. A detailed description of the aforementioned framework is presented in Algorithm 1, where $Q_{m,i}^{RSRP}$ and $Q_{m,i}^{RSRQ}$ correspond to the UE measurements in terms of RSRP and RSRQ respectively.

Algorithm 1 TS upon Connection Establishment

$$\{p\} = \emptyset$$

for $i = 1$ to number of deployed CCs do

select cell $m$ of CC $i$

$m_i = \arg\max_{i} \{Q_{m,i}^{RSRP} \}$ subject to $Q_{m,i}^{RSRP} > RSRP_{min}$ & $Q_{m,i}^{RSRQ} > RSRQ_{min}$

$\{p\} \leftarrow \{p\} \cup m_i$

end for

among the load balancing candidate set $\{p\}$ select target cell $m_i = \arg\max_{m_i} \{CAC_{m_i} \}$

initiate HO towards cell $m_i$
B. Dedicated Absolute Priorities

Absolute Priorities (AP) are a priority-based TS scheme for UEs in the RRC Idle state [6]. Carrier priorities are broadcast on the system information and devices reselect to a higher priority whenever the measured signal quality or power exceeds the \( \text{Thresh}_{X_{\text{high}}} \) threshold. On the other hand, reselections towards a lower priority CC occur only if the serving measurement falls below the \( \text{Thresh}_{X_{\text{low}}} \) threshold and the target exceeds \( \text{Thresh}_{X_{\text{low}}} \).

Nevertheless, dedicated UE to carrier priority association is also feasible by explicitly providing the device with an updated priority list during the connection release. In this case, the Dedicated Absolute Priorities (DAP) overwrite the broadcast ones. As a continuation of our previous work [12], a dynamic load-based DAP framework for HetNet deployments is illustrated in Algorithm 2. In principle, priorities are dynamically adapted based on the cell load conditions. Similarly to Algorithm 1, the DAP scheme takes into account both coverage and interference limitations as well. Nevertheless, \( \text{RSRP}_{\text{min}} \) is changed to the \( \text{Thresh}_{X_{\text{low}}} \) threshold. Therefore, by assigning the highest priority to the least loaded layer, it is ensured that the UE camps at the underutilized CC after releasing its connection and switches to RRC Idle.

**Algorithm 2** Dedicated Priorities

\[
\{p\} = \emptyset
\]

for \( i = 1 \) to \( \text{number of deployed CCs} \) do

select cell \( m \) of \( \text{CC} \ i \)

\( m_i = \arg \max_m \{Q_{m,i}^{\text{RSRP}}\} \) subject to

\( Q_{m,i}^{\text{RSRP}} > \text{Thresh}_{X_{\text{low}}} \) and \( Q_{m,i}^{\text{RSRQ}} > \text{RSRP}_{\text{min}} \)

\( \{p\} \leftarrow \{p\} \cup m_i \)

end for

sort \( \{p\} \) in descending CAC order and derive the corresponding CC list \( \{f\} \)

assign the highest priority to the first CC of set \( \{f\} \) and continue on a descending order

IV. SIMULATION ASSUMPTIONS

The joint operation of TS with CA is evaluated in scenarios A and B, as those are described in the introductory section. Finite buffer traffic is simulated and packet arrivals are modelled as a Poisson process. The burst size is negatively exponentially distributed with a mean value of 4 Mbits. For both scenarios, the mean packet interarrival time is accordingly adjusted such as the offered load corresponds to \( \sim 80\% \) of the system capacity. 2 high traffic areas are generated per site and picocells are deployed concentrically. The minimum allowed pico-to-pico distance is set to 40 m. Hotspot UEs are confined within an area of 20 m radius, while the free moving users follow straight line trajectories. Low mobility at 3 km/h is assumed for both user types. As a reference, the broadcast AP+CA case is utilized. The highest priority is assigned to the 2600 MHz layer, such as to exploit small cells as much as possible. On the other hand, TS+CA applies the proposed load balancing mechanisms in the RRC Idle and Connected. All TS operations occur solely at the PCell. SCell actions are based on RSRQ measurements. In general, a rather aggressive SCell addition policy is applied in order to fully exploit CA and improve trunking efficiency. Joint cross-CC PF scheduling is applied at the macro CCs, while conventional PF is assumed for the picocells. The key simulation assumptions are outlined in Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layout</td>
<td>7 sites, 3 sectors per site</td>
</tr>
<tr>
<td>Inter-Site Distance</td>
<td>500 m</td>
</tr>
<tr>
<td>Pathloss</td>
<td>3GPP Models</td>
</tr>
<tr>
<td>Deployed CCs (Bandwidth)</td>
<td>CC1: 2600 MHz (20 MHz)</td>
</tr>
<tr>
<td></td>
<td>CC2: 800 MHz (10 MHz)</td>
</tr>
<tr>
<td></td>
<td>CC3: 1800 MHz (10 MHz)</td>
</tr>
<tr>
<td>#picocells per sector area</td>
<td>2</td>
</tr>
<tr>
<td>#UEs per sector area</td>
<td>100 @ 3km/h</td>
</tr>
<tr>
<td>Hotspot UE Density</td>
<td>66%</td>
</tr>
<tr>
<td>Transmit Power</td>
<td>{43, 30} dBm</td>
</tr>
<tr>
<td>Antenna Configuration</td>
<td>1x2</td>
</tr>
<tr>
<td>Traffic Type</td>
<td>Finite Buffer @4Mbits</td>
</tr>
<tr>
<td>CA Packet Scheduler</td>
<td>Cross-CC PF</td>
</tr>
<tr>
<td>RRC Idle AP</td>
<td>Highest priority @2600MHz</td>
</tr>
<tr>
<td></td>
<td>( \text{Thresh}<em>{X</em>{\text{low}}} ): -110 dBm</td>
</tr>
<tr>
<td></td>
<td>( \text{Thresh}<em>{X</em>{\text{high}}} ): -106 dBm</td>
</tr>
<tr>
<td>RRC Connected</td>
<td>( \text{Intra-HO: RSRP-based (A3 event)} )</td>
</tr>
<tr>
<td></td>
<td>( \text{Intra-HO Offset: } 3\text{dB (TTT}=0.4\text{s}) )</td>
</tr>
<tr>
<td></td>
<td>( \text{Inter-HO: RSRP-based (A3 event)} )</td>
</tr>
<tr>
<td></td>
<td>( \text{Inter-HO Offset: } 4\text{dB (TTT}=0.5\text{s}) )</td>
</tr>
<tr>
<td></td>
<td>( \text{A2 Event Threshold: } -110 \text{dBm} )</td>
</tr>
<tr>
<td>SCell Events (RSRQ-based)</td>
<td>Addition: target cell above (-16) dB</td>
</tr>
<tr>
<td></td>
<td>Removal: SCell falls below (-18) dB (or UE switches to RRC Idle)</td>
</tr>
<tr>
<td></td>
<td>Change: target becomes ( 3 \text{dB better than } \text{SCell} )</td>
</tr>
<tr>
<td>Maximum SCell support</td>
<td>1 SCell</td>
</tr>
<tr>
<td>TS Framework</td>
<td>Dedicated Priorities</td>
</tr>
<tr>
<td></td>
<td>Forced HO upon Connection Setup</td>
</tr>
<tr>
<td></td>
<td>( R_{\text{target}} : 6 \text{Mbps} )</td>
</tr>
<tr>
<td></td>
<td>( { p_{\text{target}}, p_{\text{max}} } : { 0.85, 0.2 } )</td>
</tr>
</tbody>
</table>

V. SIMULATION RESULTS

Fig. 2 illustrates the PCell distribution for different CA UE ratios. Regardless of the investigated scenario, PCells are solely assigned onto the 2600 MHz layer when broadcast AP are applied. This behavior is expected since all UEs are camping on this CC due to its high priority. Therefore, whenever switching to RRC Connected, the connection is established at the same layer and SCells are assigned on the escape carriers. On the hand, TS distributes the PCell assignment via the load balancing operations that are performed in the different RRC states. An interesting observation is the fact that the PCell distribution starts converging to the AP+CA one, as the CA UE ratio goes higher. In principle, the system capacity increases with the CA UE penetration for fixed offered load conditions. The \( R_{\text{target}} \) requirement is met more easily and consequently less TS-driven actions are triggered. Finally, TS slightly offloads the picocells compared to when sole AP are applied. The reason is that specific picocells experience interference problems and they cannot meet the \( R_{\text{target}} \) bit rate. Hence, a small fraction of hotspot UEs has to be steered to the inter-frequency macrocell CCs.
The throughput gains of TS+CA over AP+CA are depicted in Fig. 3. As expected, the maximum gains derive at 0% CA UE ratio since the escape carriers are poorly utilized if broadcast AP is applied. By increasing the CA penetration, better utilization of the escape carriers is achieved as they can be exploited by CA UEs. Hence, the TS+CA over AP+CA gains gradually diminish. In principle, TS provides no benefit for CA UE ratios above 50% for scenario A. Load balancing among the macro CCs is performed via cross-CC PF scheduling, while AP in the RRC Idle guarantee the high exploitation of the picocell layer.

Regarding scenario B, we observe that TS is relevant for CA UE penetrations greater than 50%. As the number of deployed CCs is greater than the multi-CC connectivity capabilities of the CA device, the CA scheduler cannot fully resolve the load imbalances created by AP. The TS-driven distribution of PCells is still required and load balancing is provided by the joint operation of CA with TS. In this case, TS+CA primarily improves the 5%-ile UE throughput since the resource share fairness between legacy non-CA and CA users is enhanced. In particular, the corresponding gains at 50% and 60% CA UE ratio are 70% and 25% respectively. Nevertheless, TS+CA does not provide any benefit for CA UE ratios above 70%. For these cases, the performance bottleneck comes from the picocell users that are offloaded to the escape CCs.

Fig. 4 shows the average UE throughput per Base Station Technology (BTS) for Scenario B. The impact of CA penetration on the experienced throughput at the macro layer is rather noticeable for the AP+TS case. As the CA UE ratio increases, higher data rates are achieved; downlink buffers empty at a faster rate and users switch to RRC Idle. Therefore, co-channel interference at the 2600 MHz CC decreases, resulting in significant throughput gains for the picocells as well. When TS+CA is applied, picocell throughput further enhances. However, the small portion of picocell UEs that are offloaded to the escape macro CCs tend to increase their load. Compared to the AP+CA case, this effect leads to lower data rates for high CA penetrations at the macrocell layers. For CA UE ratios above 70%, the achievable picocell gains do not compensate the macrocell losses resulting in the marginal overall degradation observed in Fig. 3. Regardless of their connection capabilities, hotspot devices should be kept at the picocell layer at the cost of degraded spectral efficiency, leaving macro resources for the devices that are away from the small cell vicinity.
resulting in less TS-driven actions. Although not presented, similar trends for scenario A were observed as well.

Finally, Fig. 6 illustrates the SCell event rate for Scenario B. Compared to the corresponding HO rates, RRC signaling is dominated by the SCell events. For the AP+TS case, the overall SCell-related signaling decreases with the CA UE ratio simply because the absolute number of SCell events increases at a slower rate, compared to the number of CA users. An interesting effect is the fact that TS+CA significantly reduces SCell changes up to 50%-80%. The reason is the fact that the RSRQ distributions of the macro CCs are more balanced and therefore the SCell change condition is less likely to be met. In general, load imbalances are also reflected at the RSRQ distributions. Those are partially compensated by more SCell changes when AP+CA is applied and especially at low/medium CA UE penetrations. On the other hand, slightly higher SCell additions/removals are observed for TS+CA due to the aforementioned pico-to-macro offloading triggered by TS. Nevertheless, the overall SCell-related signaling for TS+CA is lower compared to the corresponding AP+CA one. However, the significant SCell changes gains are not reflected in the overall signaling reduction as the amount of SCell additions/removals is significantly higher due to the aggressive SCell addition policy.

VI. Conclusions

In this paper, a load-based TS framework has been developed for investigating its interaction with Release 10 Intra-Site CA. System level simulations had been conducted for different LTE-Advanced HetNet deployments and CA UE ratios. Regardless of the investigated scenario, results have shown that TS is rather critical for any CA UE ratio below 50% in order to efficiently utilize network resources. For CA penetrations above 50%, TS is still relevant in scenarios where the number of deployed CCs is greater than the number of serving CCs that a CA device supports. In these deployment cases, the load among the available CCs cannot be balanced solely by the CA scheduler and consequently TS-driven HOIs/cell reselections are required. Such approach primarily benefits the 5%-ile UE throughput as enhanced resource share fairness between legacy non-CA and CA UEs is achieved. On the other hand, peak data rates are not noticeably improved. The CA penetration breaking point, where the TS gains diminish, depends heavily on the TS configuration along with the interference experienced by the picocell layer. In the context of mobility performance, CA SCell events dominate the generated RRC signaling, being significantly higher compared to the corresponding HO rates. Nevertheless, TS can reduce the amount of SCell-related events, while the actual signaling gains depend on the aggressiveness of the SCell addition policy.

REFERENCES