QoS-Based Virtualization of User Equipment in 5G Networks

(*Invited Paper*)

Slim Zaidi*, Oussama Ben Smida*, Sofiène Affes*, Usa Vilaipornsawai†, Liqing Zhang†, and Peiying Zhu†
*INRS-EMT, Université du Québec, Montreal, QC, H5A 1K6, Canada, Email: {zaidi, oussama.ben.smida, affes}@emt.inrs.ca
†Huawei Technologies Canada Co. Ltd., Canada, E-mails: {usa.vilaipornsawai,liqing.zhang,peiying.zhu}@huawei.com

***Abstract*—This paper develops an innovative scalable and low-cost QoS-based user equipment (UE) virtualization (UEV) scheme that capitalizes on the massive connectivity and the enhanced resources of the new UE generation in terms of connectivity, computing, battery/power, etc. Exploiting the need/excess duality and the heterogeneity in such resources at the UE plan, we form virtual UEs (VUE)s, dynamically, owing to a carefully-designed time-adjusting scheme for the selection of the proper cooperative UE sets. The new UEV scheme is able to adapt to each target UE (TUE) environment, meet its demands, and scale with its needs, offers a reliable and efficient yet low-cost inter-UE cooperation, reduces the overhead and power consumption with respect to conventional approaches, and substantially reduces the number of communication links and, hence, incurs much less interference. System-level simulation results show that the proposed QoS-based UEV scheme largely outperforms the "dummy UEs" approach.**

***Index Terms*—User equipment (UE) virtualization (UEV), wireless access virtualization (WAV), 5G, user-centric architecture, cloud UE, device-to-device (D2D) communications**

I. INTRODUCTION

Current 4G radio access networks (RAN)s adopt cell-centric architectures where the cell is the network's focal point which serves several UEs located in its coverage area [1]-[5]. As the number of users and the services' data rate increase, conventional cellular networks, whose spectrum resources are limited, approach their limits. One straightforward way to cope with such a huge mobile data traffic is to increase the system capacity by deploying more and more transmission points (TP)s. This allows not only spectrum reuse across large geographic areas, but also the reduction in the number of devices competing for each TP's resources. Unfortunately, extreme densification leads inevitably to high inter-cell interference and, hence, a poor cell-edge user experience. In order to overcome this liming factor, some remedial solutions such as inter-cell interference coordination, coordinated beamforming [6], and fractional frequency reuse have been introduced in 4G RAN. Although the latter offered some performance gains at the cost of increased complexity and overhead, they were unable to completely remove the cell-boundary effect.

By capitalizing through cell virtualization on this clear trend of extreme densification, future 5G networks will provide, in contrast to their predecessors, boundaryless communications and ensure uniform and consistent user experience [7]. This would potentially lead to substantial improvements in terms of network's spectral and power efficiencies and, hence, to the fulfillment of 5G's pledge of ubiquitous user experience. On the other hand, massive connectivity is also envisioned in future 5G networks [8]-[9]. Indeed, recent years have witnessed the surging popularity of mobile services and applications, resulting in an explosive growth in the number of mobile devices and tremendous advances in their capabilities. This offers new opportunities to achieve edgeless communication through wireless access virtualization (WAV) at the UE level by inter-UEs cooperation. Indeed, multiple UEs may be grouped to form a VUE with enhanced capabilities, thereby allowing efficient resource allocation and coordination and improved network performance. Nevertheless, UEV raises several challenges mainly related to its administration and incentives, the management of the incurred interference, and the selection of cooperative UEs, their privacy and security, and their battery lifetime.

Aiming to address most of these concerns, [5] proposed to perform UEV through a network-aware device with enhanced processing and front-end capabilities. Called dummy UE, such a device uses the same air interface as any other UEs to assist their discovery and help them transmitting and receiving data. It is nothing but a relatively cheap device deployed by the network operator to enhance performance or relieve the traffic stress caused by hotspots. Indeed, by placing the so-called dummy UE at an optimal geographic position sufficiently high to give it some noticeable SINR gain (e.g., 5 dB) over the surrounding ordinary UEs, it offers not only substantial throughput and coverage gains, but also enables the implementation of advanced signal processing techniques that require accurate CSI estimates and powerful processors. However, despite its efficiency, the so-called dummy UEs approach suffers from several drawbacks. First, it requires pre-planning and, hence, is unable to handle unpredictable hotspots. Second, although dummy UEs are relatively cheap, their deployment and maintenance can be very expensive since they are envisioned to be introduced by the hundreds and placed at presumably optimal yet very often hardly accessible locations. Besides, these supposedly strategic positions might fall short from delivering the target SINR gain due to unpredictable interference caused either by neighboring personal femtocells or other in-band device-to-device (D2D) transmissions. The latter may cause dramatic performance degradation and even render these dummy UEs useless. Finally, the so-called dummay UEs approach is unscalable since it would require more dummy UEs to cope with the growth of UEs. Even worse, these dummy UEs might quickly become

---

Work supported by the NSERC/Huawei Canada/TELUS CRD Grant on 5G-WAVES (WAV Enabling Schemes), the DG and CREATE PERSWADE <www.create-perswade.ca> Programs of NSERC, and a Discovery Accelerator Supplement Award from NSERC.

978-1-5386-2070-0/18/$31.00 ©2018 IEEE

obsolete due to the rapid technological advances in hardware.

Due to the aforementioned issues, authors in [10]-[18] have opted for inter-UEs cooperation. The latter is actually a more scalable virtualization solution that enables the efficient use of available resources without requiring any costly network infrastructure improvements. A pressing and critical question arises, however, as how to select the set of cooperative UEs to rip the maximum benefits of their cooperation and significantly improve performance without increasing costs? Conventional relay selection techniques such as [10]-[18] could be adopted. However, they require huge information exchange between all UEs that translate into significant extra overhead, latency, and power consumption, all condemned to increase exponentially with the network densification and massive connectivity foreseen in future 5G networks. Furthermore, the VUEs are usually formed in these works using highly-complex iterative greedy search algorithms that explore all potential set constructions to ultimately settle on groups that are very often far from optimal. On the other hand, by activating some UEs to cooperate with one target UE (TUE) that needs help, the number of communications link increased, thereby depleting system resources and causing high interference level.

Motivated by all these issues, we develop in this paper an innovative scalable and low-cost QoS-based UEV scheme that capitalizes on the massive connectivity and the enhanced resources of the new UE generation in terms of connectivity, computing, battery/power, etc. Exploiting the need/excess duality and the heterogeneity in such resources at the UE plan, we form VUEs, dynamically, owing to a carefully-designed time-adjusting scheme for the selection of the proper cooperative UE sets. The new UEV scheme is able to adapt to each TUE environment, meet its demands, scale with its traffic needs, offers a reliable and efficient yet low-cost inter-UE cooperation, reduces the overhead and power consumption with respect to conventional approaches, and substantially reduces the number of communication links and, hence, incurs much less interference. Simulation results show that the proposed QoS-based UEV largely outperforms the ”dummy UEs” approach.

II. 5G SYSTEM MODEL

As illustrated in Fig. 1, our system consists of an ultra-dense subnetwork comprised of extremely heterogenous system infrastructure and UEs. The latter are assumed to have different radio access technologies (RAT)s, antennas numbers, processing capabilities, and, powering technologies. It is this extreme heterogeneity, especially at the UEs plan, that will characterize future 5G networks. As shown in Fig. 1, in such a network some UEs may be served by macro, micro, or femto base stations (BS)s while others are served by a set of BSs forming a virtual BS. Furthermore, some UEs may have access to private networks through a different RAT such as WiFi while others might have access to a particular network infrastructure through uncrowded bands such as mm-wave. Some UEs might not be advanced enough to benefit from it. All UEs in Fig. 1 could be co-located or relatively close to each other. The subnetwork of our concern could be any hotspot created by outdoor festivals or rallies, coffee shops, or shopping malls. It is noteworthy here that some of UEs in these hotspots might belong to the same subscriber or owner. They could be smatphones, tablets, and/or wearable devices.

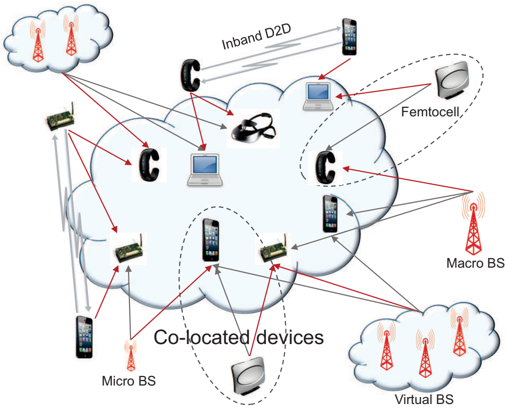


Fig. 1: System model.

Such heterogeneity in both UEs and infrastructure plans give rise to new use-cases wherein some UEs might have access to more resources than required when others would be in shortage. In such a case, the latter will be simply provided by new resources, thereby depleting the system resources. Another phenomenon that will also characterize future 5G systems and must be underlined here is the unpredictable interference due mostly to femtocells and in-band D2D. Both are extremely random and, hence, unpredictable since they are often activated by the subscriber itself in order to extend its coverage or access a wider range of services. Consequently, the network cannot predict such interference as it would do with inter- and intra-cell interference in the current conventional 4G context.

All these phenomena and 5G context characteristics must be taken into account when designing the prospective UEV strategy.

III. PROPOSED QOS-BASED UEV

Fig. 2 shows the link capacity and QoS profiles of a prospective 5G network. The 5G link capacity profile is anisotropic due to the extreme heterogeneity of both infrastructure and UE sides. Indeed, as discussed above, some co-located (or closely located) UEs may access different quantities of resources since they have different available RATs or priorities. They could also suffer from different levels of unpredictable low-range interference. All these factors should significantly broaden the differences between their link capacities even if they are very close to each other. This is in contrast with 4G networks wherein co-located UEs may access the same network and are subject to almost the same interference level, making its link capacity profile totally isotropic. Actually, this is one of the main conceptual evolutions that should govern 5G standards. Besides, as illustrated in Fig. 2,

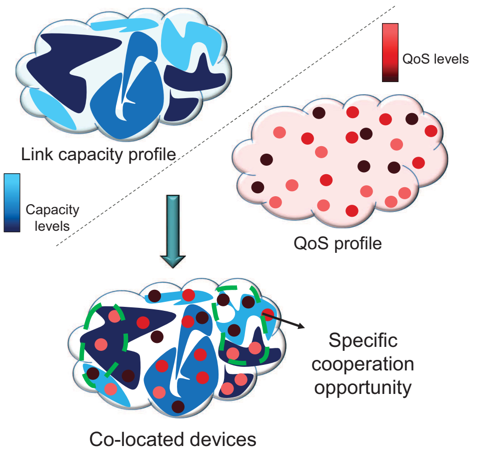


Fig. 2: 5G link capacity and QoS profiles.

the QoS requirements of co-located UEs considerably varies from an UE to another. This is a valid assumption since 5G networks will support a wider range of services and applications with much different QoS requirements. Actually, this assumption is already valid in 4G context wherein for instance UEs in the same vicinity of say a coffee shop could launch different applications such as HD video streaming web browsing, or simple file downloading with different QoS requirements.

The link capacity and QoS profiles are combined in Fig. 2 to reveal many new cooperation opportunities that should be exploited in the prospective UEV approach. Indeed, some of the dark red discs standing for UEs with high QoS requirement are located in light blue regions representing poor link capacities. Whereas, UEs requiring low QoS (i.e., light red discs) would have access to much better link capacities. Motivated by the aforementioned observations, we introduce in this work a novel QoS-based UEV scheme that relies on the opportunistic association of high-QoS UEs having poor links with low-QoS UEs having strong links. In what follows, we will show how the former could significantly benefit from the latter to balance their QoS deficits with the others' excesses. We will also show how future 5G networks could benefit from our new UEV as well.

Two scenarios are considered here: i) Scenario 1 wherein some UEs have additional unused resources provided through private subnetworks or by special radio access technologies (RAT)s not commonly used in the network, and ii) Scenario 2 wherein none of the UEs have extra resources.

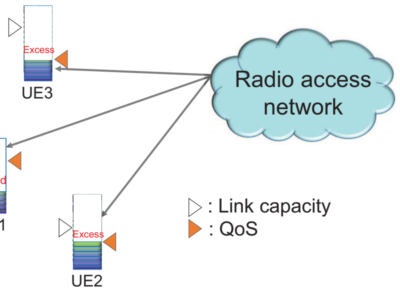
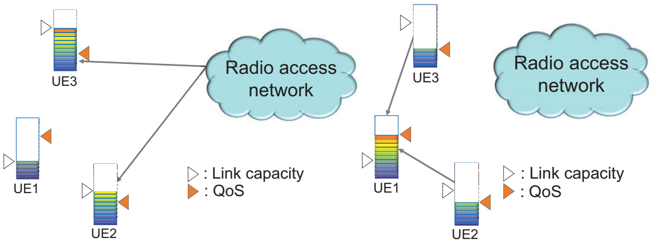


Fig. 3: Scenario 1: some UEs with extra resources.



(a) First step/hop (b) Second step/hop

Fig. 4: Proposed QoS-based UEV approach in Scenario 1.

A. Scenario 1

Fig. 4 illustrates Scenario 1 wherein three active UEs receive data during the same time slot. UE1 needs additional network resources (for example bandwidth) to cater its high QoS demands (for example throughput) while UE2 and UE3 require much less QoS but have access to extra resources unreachable by UE1 (i.e., the network is unable to serve UE1 using these extra resources due to their private character and/or limited RAT at UE1). For instance, UE1 may be unauthorized to access some surrounding WiFi networks or private femtocells and/or might not be equipped with the proper RAT to access the resources available at some BSs. So far, this issue is inefficiently addressed by providing UE1 with new network resources, if ever available, to meet its QoS requirement. Even D2D capabilities cannot be exploited to relay UE1 data through UE2 and UE3 since all three UEs are scheduled at the same time slot. Rather, this work sees in this situation a huge potential for UEs cooperation through resource sharing among UE1, UE2, and UE3. Indeed, as shown in Figs. 4(a) and 4(b), by associating these three UEs and enabling their cooperation within a VUE, the network can transmit UE1's needed data through the extra resources of UE2 and UE3 that the network could allocate to them in excess of their QoS levels up to their cellular link capacities (i.e., first step/hop) before their transfer through the D2D links to UE1 (i.e., second step/hop). The new UEV scheme actually requires a fundamental change with respect to conventional RAN, since the data sent to UE2 and UE3 contains not only their own information, but also that of UE1. It could be implemented by inserting some additional pilot symbols fractions of a given data frame with the UEs they are destined to. Nevertheless, such concept certainly requires enhanced UE capabilities as already envisioned in 5G, as well as a fundamental review of the data frame's structure that would rethink, among other things, pilot insertion rates and positions. By selecting and allowing the cooperation of active (i.e., scheduled) UEs based on their QoS (i.e., associating high-QoS UEs having less resources with low-QoS UEs having more resources), we open the door to an exponentially increasing number of cooperation opportunities. Furthermore, not only would UE1's QoS requirements be met, but any depletion of network resources would be avoided thereby resulting in substantial throughput or spectrum efficiency improvements. All these benefits highlight the efficiency at low cost of the proposed UEV approach, making it a very suitable candidate for WAV at the UE level in future 5G networks.

B. Scenario 2

In Scenario 2, none of the three scheduled UEs has extra resources. The network accounts for the link quality (measured through the SINR, CSI, or CQI, etc.) to allocate to the UEs only the amount of resources required to meet their QoS demands, as shown in Fig. 5. When UE1 increases its demand, the network may provide it with more resources (for example bandwidth) in order to meet its new requirements, as it is done in current RANs and illustrated in Fig. 6. Since UE1 is experiencing poor link conditions, large network resources must be allocated to it, thereby depleting its limited resources. In order to circumvent this serious issue, we propose in this paper that the resources be allocated to UE2 and/or UE3 instead of UE1 as shown in Fig. 7. Since UE2 and UE3 are subject to much better link conditions, much less resources are required to meet the additional UE1 demands. Again here the data sent from UE2 and UE3 contains not only their own information but also that of UE1. According to LTE standards, up to 44 times less bandwidth could be used if the proposed UEV is implemented to group high-QoS/poor-link UEs with lower QoS/better-link UEs [19]. This proves that the proposed UEV approach may provide substantial gains even when UEs have the same RATs and access the same network resources. The selection must be done here by accounting for the link condition rather than the link capacity as in Scenario 1. We will see in the next Section that it is even possible to make a UE cooperate with others having the same and different RATs and accessing the same and different private resources. This highlights again the efficiency of the proposed approach and its potential to increase the cooperation opportunities, thereby capitalizing on the massive connectivity foreseen in 5G.

Although Scenarios 1 and 2 underline two particular use-cases wherein UEs have access to different or same private networks and special RATs, they both fall under the same roof of resource sharing between UEs, that is by capitalizing on their diversity, high density, and increased smart capabilities all envisioned in future 5G.

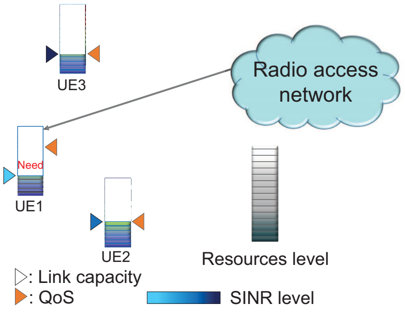


Fig. 5: Scenario 2: no extra resources.

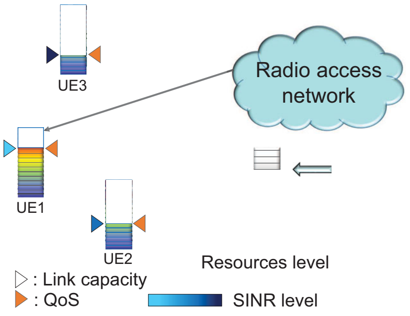


Fig. 6: Conventional approach in proposed QoS-based UEV approach in Scenario 2.

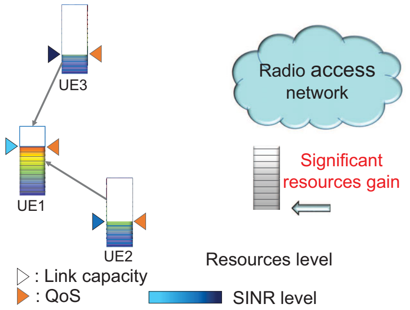


Fig. 7: Proposed approach in proposed QoS-based UEV approach in Scenario 2.

IV. ENABLING MECHANISM OF QOS-BASED UEV

This section investigates mechanisms that enable the proposed QoS-based UEV scheme. Let us first consider a network of U UEs. Let c_u [bits/s/Hz] and q_u [bits/s] denotes the u-th UE channel capacity and its required QoS, respectively. As in any orthogonal frequency division multiple access (OFDMA) air interface[1], the scheduler assigns at a given time slot t the available subcarriers (i.e., spectral resources) to K UEs among U so that a balance between predefined individual and/or global goals is reached. Let w_u denote the bandwidth allocated by the scheduler to the u-th UE and w_u^{ad} its additional unused bandwidth provided by a private subnetwork (ex., private femtocell, WiFi network, etc.) it might access or by some other special RATs not commonly used in the network.

The K active (i.e., scheduled) UEs can be then divided into two sets: the set of *in-need* UEs (nUE)s which need more resources to cater their target throughput

$$S_{\text{nUE}} = \{i|c_i w_i < q_i\}, \tag{1}$$

where $w_i^{\text{Ad}} = 0$, $i \in S_{\text{nUE}}$ since nUEs do not have any additional bandwidth as they are "*in-need*", and the set of

---

[1] OFDMA is considered here only as one exemple. our approach easily extends and applies to to any other RAT such as time-division multiple access (TDMA), code-division multiple access (CDMA), sparse code multiple access (SCMA), non-orthogonal multiple access (NOMA) , etc, or any combination thereof.

Algorithm 1 Proposed QoS-based UEV algorithm.

```
Input: nUE, pUE, $c_u$, $w_u$, $w_u^{ad}$, $Ne_u$, $q_u$, $R_{i,u}$;
$S_{\text{nUE}}$ = $\{i;\ c_i w_i < q_i\}$;
$S_{\text{pUE}}$ = $\{p;\ c_p(w_p + w_p^{ad}) \geqslant q_p\}$;
for i $\in S_{\text{nUE}}$ do
  $V_i$ = $\emptyset$;
  $C_{\text{pUE}}^i$ = $\{k \in S_{\text{pUE}};\ R_{i,k} \geqslant q_i - c_i w_i\}$
  $k_0$ = $\underset{k \in C_{\text{pUE}}^i}{\arg\max}(c_k)$;
  if $c_{k_0} \geqslant c_i$ then
    $V_i$ = $V_i \cup k_0$;
    if $w_i c_{k_0} > Ne_i$ then
      $w_{k_0}^{ad}$ = $\frac{w_i c_{k_0} - Ne_i}{c_{k_0}}$ + $w_{k_0}^{ad}$;
    else
      $Ne_i = Ne_i - w_i c_{k_0}$;
    end if
  end if
  while $Ne_i > 0$ or $C_{\text{pUE}}^i \neq \emptyset$ do
    $k_0$ = $\underset{k \in C_{\text{pUE}}^i}{\arg\max}(w_k^{ad} c_k\ )$;
    $V_i$ = $V_i \cup k_0$;
    $C_{\text{pUE}}^i$ = $C_{\text{pUE}}^i \backslash \{k_0\}$;
    if $w_{k_0}^{ad} c_{k_0} \geqslant Ne_i$ then
      $Ne_i$ = 0;
      $w_{k_0}^{ad}$ = $w_{k_0}^{ad}$ - $\frac{Ne_i}{c_{k_0}}$;
    else
      $Ne_i$ = $Ne_i$ - $w_{k_0} c_{k_0}$;
      $w_{k_0}^{ad}$ = 0;
    end if
  end while
end for
```

provider UEs (pUE)s able to provide nUEs with data thanks to their additional resources and/or high channel capacity

$$S_{\text{pUE}} = \left\{k | c_k(w_k + w_k^{ad}) \geq q_k\right\}. \tag{2}$$

In the sequel, we propose an algorithm that selects the best cooperative pUEs forming with nUEs VUEs V_is wherein nUEs meet their required QoS using the available network resources. As a first step, we need to determine for each nUE$_i$ a set of cooperative candidates C_{pUE}^i. To this end, nUE$_i$ selects from all neighboring UEs that it can have a D2D communication with any UE$_i$ that is potentially able to serve it by verifying if that D2D link satisfies

$$R_{i,u} \geq Ne_i = q_i - c_i w_i, \tag{3}$$

where $R_{i,u}$ [bits/s] is the maximum achievable rate of the D2D link and Ne_i [bits/s] is the needed data rate at nUE$_i$ to meet its QoS. Note that nUE$_i$ selects its candidates from the whole set of active UEs in its vicinity since it is oblivious to their allocated bandwidth and required QoS and, hence, to S_{pUE}. It is also noteworthy that the condition in (3) aims at reducing the number of cooperative UEs to 1, in order to decrease the number of D2D links. Indeed, in such a case, a UE having enough additional resources to cover the need Ne_i can relay it alone the D2D link and, hence, no other pUE is required. nUEi feeds back its subset of UE candidates to the network and the latter refines and reduces it by keeping only the pUEs willing to collaborate to the following cooperative subset of pUE candidates:

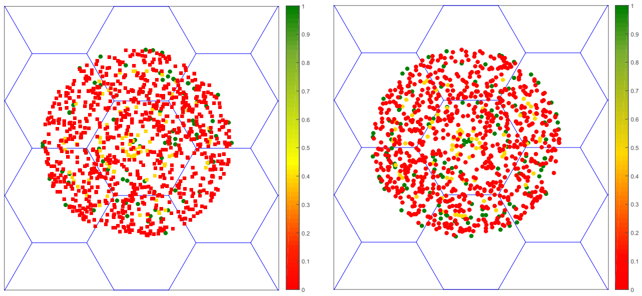
$$C_{\text{pUE}}^i = \{k | R_{k,i} \geq Ne_i = q_k - c_n w_k\}. \tag{4}$$

In order to optimize its resources usage, one should start allocating the bandwidth of nUE$_i$ to a pUE$\in$ C_{pUE}^i that has a better channel capacity before moving on to exploiting its additional bandwidth. This would not only increase the spectral efficiency (i.e., more delivered data using the same resources), but would also enhance the cooperation opportunities of the other nUEs in need and waiting for their turn to be served as well. If several candidates with better channel conditions exist, we should naturally pick the best one (i.e., $k_0 = \arg\max_{k \in C_{\text{pUE}}^i} \{c_k\}$), allocate UE$_{k_0}$ to V_i, and check whether $w_n c_{k_0} \geq Ne_i$. If this condition is satisfied, then UE$_{k_0}$ is able to satisfy the required QoS of nUE$_i$ and, therefore, its additional bandwidth is updated as follows

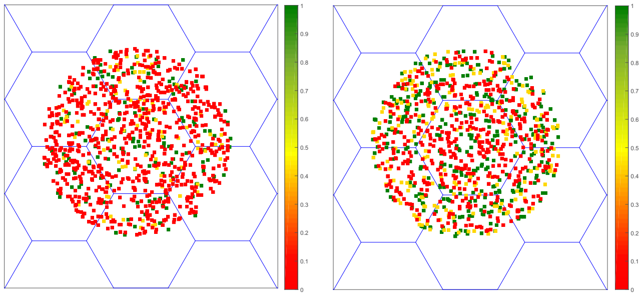
$$w_{k_0}^{ad} = w_{k_0}^{ad} + \frac{w_i c_{k_0} - Ne_i}{c_{k_0}}, \tag{5}$$

where the second term of the right-hand-side (RHS) of the equation above stands for the remaining bandwidth (if any) after providing nUE$_i$ with sufficient data to meet its QoS. Should we have $w_i c_{k_0} < Ne_i$, then and only would we explore the possible exploitation of any additional bandwidth resources pUEs $\in$ C_{pUE}^i might have to provide nUE$_i$ with $Ne_i = Ne_i - w_i c_{k_0}$. Again here we propose to select pUE$_k$ that maximizes $w_k^{ad} c_k$ in order to leave as many cooperation opportunities as possible to the other nUEs. Once this best candidate is found, the network allocates it to V_i, then checks its ability to entirely deliver Ne_i. If so, its additional bandwidth is reduced by the spectrum amount required to cover Ne_i. Otherwise, all its additional bandwidth is fully exploited to partially cover Ne_i. Then we keep moving on to the next best remaining pUE$_k$ and repeat the last steps above until $Ne_i = 0$ (i.e., the nUE$_i$'s QoS is satisfied) or the set C_{pUE}^i is totally exhausted. Algorithm 1 summarizes all the UE selection and VUE formation steps of the proposed QoS-based UEV scheme discussed above.

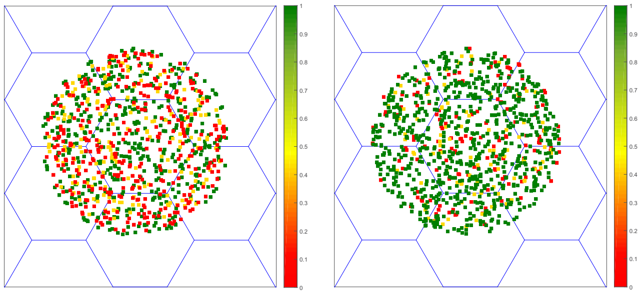
By exploiting the QoS dimension as shown in (2)-(5), our proposed QoS-based UEV approach encompasses various cooperation scenarios resulting from the heterogeneity of both UE and infrastructure plans characterizing future 5G networks. Indeed, UEs with the same channel conditions (bad or good) could help each other should they have different QoS requirements. Besides, UEs with relatively good channel conditions and low QoS may help others with much worse channel conditions and higher QoS requirements. This is in sharp contrast with existing techniques that consider a sole scenario wherein high-capacity UEs assist those with low capacity [10]-[18]. Section VI will show that the significantly increased cooperation opportunities allowed by our QoS-based UEV scheme offers substantial improvements both in network throughput and spectral efficiency and also in terms of required QoS satisfaction.



(a) No D2D communications (b) 100 dummy UEs at 1 dB



(c) 200 dummy UEs at 1 dB (d) 100 dummy UEs at 5 dB



(e) 200 dummy UEs at 5 dB (f) Proposed QoS-based UEV

Fig. 8: UE satisfaction level.

V. BENEFITS OF PROPOSED QOS-BASED UEV

The benefits of the proposed QoS-based UEV approach are discussed and summarized below.

- **Dynamic and adaptive:** The proposed approach selects an appropriate set of cooperative pUEs able to help nUE meet its QoS requirement. This set is dynamically adjusted to cope with the variations of nUE's operating conditions and environments (i.e., its QoS demands, its link capacity/quality, the number of potential cooperative UEs, their link conditions and QoS requirements, etc.).
- **Low complexity:** By exploiting the QoS dimension, the proposed approach substantially simplifies the selection process of cooperative UEs. Indeed, in contrast with conventional approaches, it avoids resorting to complex yet sub-optimal iterative greedy algorithms that explore all potential set constructions to ultimately settle on groups that are very often far from optimal.
- **Low overhead and power costs:** The selection process can be implemented locally, at the TUEs, thereby avoiding the prohibitive overhead and power costs that would have been incurred if we were to proceed in a conventional way by feeding back all CSI to the network.
- **Reduced interference level:** The association of high-QoS/low-capacity UEs with relatively lower-QoS/higher-capacity UEs opens the door towards a dramatic reduction in the number of UE-to-network (or user-to-infrastructure) communication links and, hence, in the network's interference levels. Indeed, a VUE may transmit/receive data and signaling[2] only through small UEs' subsets including, thereby decreasing communication links' number. This is in contrast with conventional UEs selection algorithms which activates non-active UEs to make them cooperate with TUE and, hence, increases even more the communication links and their incurred interference [10]-[18]. This is actually another conceptual difference characterizing our approach using which TUEs data is transmitted along with cooperative UEs data.
- **Improved network resource management:** The proposed UEV approach allows tremendous savings in network resources. Indeed, not only it significantly reduces the number of communication links, but it also keeps only the best ones that require much less resources to achieve exactly the very same target performance. As mentioned in Section III, according to the LTE standard, our approach would require until 44 times less resources in the spectral domain [19].
- **Scalability:** The more UEs we have in the network, the more cooperation opportunities arise for even better exploitation by our QoS-based UEV scheme, hence, the higher are its performance and added value. Consequently, our approach may capitalize on the massive connectivity foreseen in 5G networks. This is in contrast with the so-called "dummy UEs" approach [5] whose complexity and cost explode with the number of UEs.

VI. SIMULATION RESULTS

In this section, we verify through system-level simulations the efficiency of the proposed QoS-based UEV scheme and compare its performance with the "dummy UEs" approach [5] and the "best relay selection" strategy wherein each UE selects the best non-active (i.e., non-scheduled) UE that could relay data to it at the highest rate [12]. We consider an heterogenous network of 7 hexagonal macros with inter-site distance of 1000 m (i.e., 500 m radius) and transmit power of 46 dBm and 200 uniformly distributed picos with 100 m radius and transmit power of 23 dBm. We assume that each cell sector/site (i.e., macro/pico) and each UE has two transmit and two receive antennas, respectively. Each macro and all

---

[2] It is noteworthy that different UE subsets can be considered for signaling and useful data communication.

Tab. I: Achieved UE data rate and satisfaction level with 100 and 200 picos for $(q_1, q_2) = (0.5, 1.5)$ Mbps.

	100 picos				200 picos			
	Average rate [Kbps]	5%-ile [Kbps]	Full satisfaction [%]	Average satisfaction	Average rate [Kbps]	5%-ile [Kbps]	Full satisfaction [%]	Average satisfaction
No D2D	507.80	109.20	5.70	0.187	587.20	120.00	6.50	0.220
Best relay	589.90	111.60	8.20	0.214	736.80	177.40	21.70	0.422
Dummy UEs at 1 dB	612.30	123.70	14.20	0.327	668.30	145.90	18.20	0.367
Dummy UEs at 5 dB	1,180.30	308.90	43.70	0.539	1,260.60	358.10	48.90	0.621
QoS-based UEV	1,260.60	340.20	60.03	0.742	1,450.70	517.40	80.16	0.852

Tab. II: Achieved UE data rate and satisfaction level with 100 and 200 picos for $(q_1, q_2) = (0.5, 1)$ Mbps.

	100 picos				200 picos			
	Average rate [Kbps]	5%-ile [Kbps]	Full satisfaction [%]	Average satisfaction	Average rate [Kbps]	5%-ile [Kbps]	Full satisfaction [%]	Average satisfaction
No D2D	499.30	125.90	6.20	0.207	563.50	125.60	8.50	0.317
Best relay	562.40	127.30	7.00	0.294	731.80	181.90	24.60	0.483
Dummy UEs at 1 dB	608.60	130.00	16.40	0.380	661.70	152.80	20.00	0.405
Dummy UEs at 5 dB	1,156.20	319.70	45.40	0.581	1,251.10	377.70	49.30	0.692
QoS-based UEV	1,245.03	655.00	89.53	0.950	1,382.00	812.94	96.43	0.986

picos located in its coverage area form a hyper-cell that serves the UEs using dynamic point selection. A hotspot is emulated by randomly deploying in a disc of 1000 m radius 1000 mobile UEs with nomadic/pedestrian speed of 3 km/h.

In all simulations, we consider that all UEs have access to all network's resources, except in Tab. II where we add in each macro 100 private picos whose resources are accessible only by one half of the UEs regardless of their QoS requirements. Furthermore, we assume that half of the UEs require a minimum QoS rate of q_1 Mbps versus $q_2 < q_1$ Mbps by the other half. Please note here that a user with QoS $q_1 > q_2$ whose link capacity allows it to get the data rate $R_1 > q_1$ (i.e., in excess) could transfer a portion of it, say δ_R, to a neighboring user with QoS q_2 whose link capacity gives it a data rate $R_2 < q_2$ (i.e., in need) such that $R_2 + \delta_R = q_2$ (i.e., resource transfer and balance), and yet still have as one possibility $R_1 - \delta_R > q_1$ (i.e., still in excess of minimum QoS). Hence, the achieved average rate could exceed the average target QoS level, here of $(q_1 + q_2)/2$.

For the sake of fairness, we adopt the same decode-and-forward (DF) cooperation mode both in [5] and the proposed QoS-based UEV approach. In order to compare the two schemes, we opt for the following metric:

$$\mathcal{S}_{\mathrm{UE}} = min\left(\frac{R_A}{R_T}, 1\right), \tag{6}$$

where $\mathcal{S}_{\mathrm{UE}}$ stands for the satisfaction level, R_A is the actual (i.e, delivered) rate, and R_T is the target rate (i.e., $R_T = q_1$ or $R_T = q_2$). $\mathcal{S}_{\mathrm{UE}}$ reaches its maximum 1 when the UE is fully satisfied [i.e., $R_A \geq q_1$, hence its minimum with 1 in (6) if still in excess, or $R_A = q_2$] and its minimum 0 when the latter is not served at all.

In a preliminary qualitative assessment, we plot in Fig. 8 the UE satisfaction level achieved either without applying any UEV approach or using the "dummy UEs" and the proposed UEV schemes. In particular, Figs. 8(d) and 8(e) consider 100 and 200 dummy UEs, respectively, strategically placed at optimal positions presumably providing 5 dB SINR gains over their neighboring UEs, whereas Figs. 8(b) and 8(c) reduce the latter to 1 dB due to the stronger-than-expected interference levels that could easily arise anytime from all unpredictable inband D2D or private small cell transmissions.

The conventional case of Fig. 8(a) where none of the UEV approaches is adopted, the network performs very poorly with the red-colored very-low-satisfaction level dominating. This is hardly surprising because conventional networks would be currently unable to accommodate the huge traffic demand arising from the very likely 5G hot-spot scenario considered herein. In Figs. 8(d) and 8(e), however, the "dummy UEs" approach provides noticeable UE satisfaction gains that increase (see green-colored high-satisfaction-level dots becoming more and more numerous) both with more dummy UEs (Figs. 8(b) and 8(d)) and/or higher-SINR-gain dummy-UE positions (Figs. 8(c) and 8(e)), actually more so with the latter. Indeed, the "dummy UEs" approach is extremely vulnerable to the unpredictable yet inevitable interference that reduces the SINR gains expected at the dummy-UE positions from their nominal to practical levels set here as one example to 5 and 1 dB, respectively. As such, we see from Figs. 8(c) and 8(d) that the dummy UEs approach performs better with higher SINR gains at the dummy-UE positions (i.e., 5 versus 1 dB) yet with less dummy UEs (i.e., 100 versus 200). Fig. 8(f) illustrates the performance of the proposed QoS-based UEV approach in terms of UE satisfaction level. The dominating green-colored very-high-satisfaction-level dots reveal unambiguously its superiority over the "dummy UEs" scheme in all four considered scenarios of Fig. 8(b) to 8(e), i.e., even in the most favorable conditions of a larger number of dummy UEs all placed at optimal positions offering the full 5 dB SINR gain.

Tabs. I and II quantitatively evaluate at two target QoS couples (q_1, q_2) of $(0.5, 1.5)$ and $(0.5, 1)$ Mbps, respectively, the performance of a conventional network without D2D communications, the "best relay selection" technique, and both the proposed QoS-based UEV and the "dummy UEs" schemes in terms of the average rate, the 5%-ile rate (i.e., coverage), the full-satisfaction percentage (i.e., relative number of UEs achieving $\mathcal{S}_{UE} = 1$), and the average satisfaction score (i.e., expectation of $\mathcal{S}_{UE}$ over all UEs) achieved by the UEs in the presence of 100 or 200 picos. From these tables, we report the following main observations:

- All tested schemes see their data rate performance improve, either on average or at 5%-ile, when the number of picos and/or the average target QoS rate $(q_1 + q_2)/2$ increase(s), i.e., when there is a stronger offer of and/or demand for more resources.
- All tested schemes still see their satisfaction performance improve, either in full-achievement percentage or on average, when the number of picos increases, i.e., when there is a stronger offer of resources. However, they see it decrease when the average target QoS rate $(q_1 + q_2)/2$ increases, i.e., when there is a stronger demand for more resources, thereby underlying the resulting penalty in terms of decreasing fairness among UEs.
- The "best relay selection" scheme offers some performance gains against the conventional "no D2D" case in terms of all measured metrics. However, the "dummy UEs" approach could offer much more noticeable performance gains (except in terms of rate performance at 1 dB dummy positions in Tab. II with 200 picos), more so when the dummy UEs are indeed placed optimally at higher SINR-level positions. These performance improvements remain, however, dependent on and hence vulnerable to achieving the higher SINR levels required at the dummy positions in real-world conditions. As such, they offer little guarantees in practice due to the stronger-than-expected interference levels that could easily arise anytime from all unpredictable inband D2D or private small cell transmissions. The "dummy UEs" approach can lose up to about 50% in all measured performance metrics when the dummy positions see their SINR gains fall from 5 to 1 dB.
- Most importantly, the proposed QoS-based UEV scheme unambiguously surpasses all tested benchmarks and outperforms the "dummy UEs" at 5 and 1 dB dummy UE positions, respectively, by as much as over 10% and 110% gains in terms of throughput, over 10% and 180% gains in terms of coverage (i.e., 5%-ile rate), over 40% and 320% gains in terms of full satisfaction level, and 40% and 130% in terms of average satisfaction score, respectively.

These remarkable performance gains highlight unequivocally the net superiority of the proposed QoS-based UEV approach and it perfect suitability as a candidate for WAV at the UE level in future 5G networks.

VII. CONCLUSION

This work developed an innovative scalable and low-cost QoS-based UEV scheme that forms VUEs, dynamically, owing to a carefully-designed time-adjusting scheme for the selection of the proper cooperative UE sets. The new UEV scheme is able to adapt to each TUE environment, meet its demands, and scale with its traffic needs, offers a reliable and efficient yet low-cost inter-UE cooperation, reduces the overhead and power consumption with respect to conventional approaches, and substantially reduces the number of communication links and, hence, incurs much less interference. System-level simulation results show that the proposed QoS-based UEV scheme largely outperforms the "dummy UEs" approach.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?," *IEEE J. Select. Areas Commun.*, vol. 32, no.6, pp. 1065-1082, June 2014.

[2] S. Zaidi, S. Affes, U. Vilaipornsawa, L. Zhang, and P. Zhu, "Wireless access virtualization strategies for future user-centric 5G networks," *Proc. IEEE GLOBECOM'2016*, Washington DC, USA, December 4-8, 2016.

[3] S. Zaidi, B. Hmidet, S. Affes, U. Vilaipornsawa, L. Zhang, and P. Zhu, "User-centric wireless access virtualization strategies for future 5G networks," *Proc. IEEE ICUWB'2016*, Nanjing, China, Oct. 16-19, 2016.

[4] S. Zaidi, M. Azzakhmam, S. Affes, C. Despins, K. Zarifi, and P. Zhu, "Progressive hybrid greyfield wireless access virtualization: graph-optimized dynamic utility tradeoffs between cloud, fog, and legacy RANs," *Proc. IWCMC 2017*, Valencia, Spain, June 26-30, 2017.

[5] H. Baligh, K. Zarifi, J. Ma, P. Zhu, A. Maaref, and M. Salem "Dummy users: network-aware cooperative terminals in wireless access networks," *Proc. IEEE PIMRC'2015,* Hong Kong , China, Aug. 30-Sep. 2, 2015.

[6] S. Zaidi and S. Affes, "Distributed collaborative beamforming design for maximized throughput in interfered and scattered environments," *IEEE Trans. Commun.,* vol. 63, no.12, pp. 4905-4919, Oct. 2015.

[7] K. Zarifi, H. Baligh, J. Ma, M. Salem, and A. Maaref, "Radio access virtualization: cell follows user," *Proc. IEEE PIMRC'2014,* Washington DC, USA, Sep. 2-5, 2014.

[8] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "Network function virtualization in 5G," *IEEE Commun. Mag.*, vol. 54, no. 4, pp. 84-91, Apr. 2016.

[9] M. N. Tehrani, M. Uysal, and H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 86-92, May 2014.

[10] A. Maaref, J. Ma, M. Salem, H. Baligh, and K. Zarifi, "Device-centric radio access virtualization for 5G Networks," *Proc. IEEE GLOBECOM'2014,* Austin, USA, Dec. 8-12, 2014.

[11] A. Moubayed, A. Shami, and H. Lutfiyya, "Wireless resource virtualization with device-to-device communication underlaying LTE network," *IEEE Trans. Broadcasting*, vol. 61, no. 4, pp. 734-740, Dec. 2015.

[12] N. Nomikos, D. N. Skoutas, and P. Makris, "Relay selection in 5G networks," *Proc. IWCMC 2014,* Nicosia, Cyprus, Aug. 4-8, 2014.

[13] P. K. Mishra, S. Pandey, and S. K. Biswash, "Efficient resource management by exploiting D2D communication for 5G networks," *IEEE Access*, vol. 4, pp. 9910-9922, Sep. 2016.

[14] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 136-144, June 2014.

[15] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Trans. Commun.,* vol. 61, no. 8, pp. 3541-3551, Aug. 2013.

[16] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng, and B. Jiao, "Efficiency resource allocation for device-to-device underlay communication systems: a reverse iterative combinatorial auction based approach," *IEEE J. Select. Areas Commun.,* vol. 31, no. 9, pp. 348-358, Sep. 2013.

[17] D. Feng, L. Lu, Y. Yi, G. Y. Li, G. Feng, and S. Li, "QoS-aware resource allocation for device-to-device communications with channel uncertainty," *IEEE Trans. Vehic. Technol.,* vol. 65, no. 8, pp. 348-358, Aug. 2016.

[18] J. Hao, H. Zhang, L. Song, and Z. Han, "Graph-based resource allocation for device-to-device communications aided cellular network," *IEEE/CIC ICCC'2014,* Shanghai, China, Oct. 13-15, 2014.

[19] 3GPP, "3GPP TS 36.213 V9.2.0: Evolved universal terrestrial radio access network (E-UTRA); physical layer procedures," June 2010.