Optimized Mobile Connectivity for Bandwidth-Hungry, Delay-Tolerant Cloud Services toward 5G

Osman N. C. Yilmaz[1, 2], Carl Wijting[1], Petteri Lundén[1], Jyri Hämäläinen[2]
[1] Nokia Research Center, [2] Aalto University
Espoo, Finland
Osman.Yilmaz@aalto.fi, Carl.Wijting@nokia.com, Petteri.Lunden@nokia.com, Jyri.Hamalainen@aalto.fi

***Abstract*— The availability of different generations of radio access technologies (e.g., from 2G to 4G) and multiple coverage layers (e.g., macro- and small cell-layers) results in a larger selection of connectivity options at the disposal of mobile devices, each with their own connection characteristics (e.g., in terms of quality of service, power consumption and billing). Toward heterogeneous 5G radio access networks it is essential to provide new requirements and solutions for mobile cloud connectivity in order to optimize the trade-off between the available connectivity options and timing of the communication. In this paper, we propose context-aware connectivity management for bandwidth-hungry, delay-tolerant cloud services; and address the potential gains and challenges with the help of system-level performance evaluations.**

Keywords*— *Cloud connectivity, 5G, Heterogeneous networks, Delay-tolerant services, Context-awareness, Self-organization

I. INTRODUCTION

To satisfy the needs of 2020 wireless communications society, the next generation radio communication system, i.e., the fifth generation (5G), has to be significantly more efficient and scalable than the fourth generation (4G) in terms of energy, cost and spectral efficiency [1]. Furthermore, the large diversity of use cases and the various requirements in a wide range of emerging services, i.e., from narrow-band, delay-sensitive services (low-latency machine-type communications) to bandwidth-hungry, delay-tolerant services, have to be supported [2].

In Long Term Evolution (LTE) and 4G LTE-Advanced networks, the user equipment (UE) data is scheduled based on the instantaneous channel quality measurements so that the base station (BS) can exploit the multi-user channel diversity to optimize the radio resource utilization [3]. In addition to the radio resource scheduling mechanism taking place within a single cell, current networks are able to balance the network load between different cells and radio access technologies using the radio link quality and traffic indicators [4].

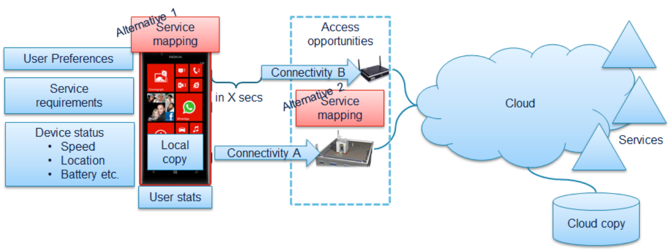

Fig. 1. Device-to-cloud service connectivity scenarios in HetNets. Alternatives 1 and 2 denote the decision points where the context information can be utilized for optimized mobile cloud connectivity.

Future mobile devices and networks are designed to learn and decide how to manage the connectivity with the assistance of context information in addition to the channel measurements and traffic indicators [5]. Context information could be locally obtained (e.g., through sensors) or fetched from the other nodes (e.g., BS and cloud servers). Context-awareness can assist UE (Alternative 1) or network (Alternative 2), as shown in Figure 1, in order to optimize the traffic management policies; and to minimize the energy-consumption of the connected nodes.

In 2020 wireless communications society, due to the increasing heterogeneity (e.g., dense small cell deployments) in cellular networks as well as new technologies (e.g., band combinations and duplexing modes), mobile devices will have a larger range of connectivity options at their disposal, which are of different characteristics in terms of quality of service (QoS), power consumption, billing and so on and so forth. To take advantage of alternative radio access opportunities in the vicinity of users, mobility-context (e.g., user velocity and proximity to the small cells) and traffic-context (e.g., service profile and network load) should be taken into account. Thus, bandwidth-hungry, delay-tolerant services, such as cloud data access and storage, can be mapped to the small cell-layer more effectively.

In this paper we first study the impact of network densification with the support of system-level performance evaluations. Following the preliminary results we introduce the context-aware cloud connectivity management; and address the potential gains and challenges of the concept in future heterogeneous networks (HetNets). Next, a complementary solution for potential synchronization conflicts is proposed. Finally, the main conclusions are given and the potential future work is discussed.

II. PRELIMINARY STUDIES

A. *Network Densification Scenario toward 5G*

Preliminary studies are carried out assuming the continuous densification of cellular networks toward 2020, i.e., gradually up to 10 times more dense deployment than in today's networks [1]. Therefore, the network layout is modeled for three network densities: 4, 12 and 20 small cells per macro sector (SC/macro), representing different future scenarios. Small cells are randomly and uniformly positioned with the minimum distance of 40 m (among each other) as depicted in Figure 2. It is assumed that macro and small cells operate at different frequency layers, for instance, while macro cells (for ubiquitous coverage) operate at the traditional LTE bands, small cells (for capacity extension) can operate at the new spectrum regimes or locally authorized spectrum bands.

The main simulation assumptions follow the simulation guidelines recommended by 3GPP [6] and the related parameters and models given in Table 1. In the scenario, UEs are randomly and uniformly dropped throughout the macro geographical area. UEs move in random directions, with 6 km/h velocity; and the direction is updated at the layout border.

TABLE I. MAIN SIMULATION ASSUMPTIONS

Network layout	3 macro cells, {12, 36, 60} small cells
Min. inter-SC distance	40 m
Carrier frequency	2 GHz
Carrier bandwidth	10 MHz (Uplink), 10 MHz (Downlink)
Downlink max. tx power	MC: 46 dBm, SC: 30 dBm
Uplink power control [7]	Pmax:21 dBm, P0: -54 dBm, alpha: 0.6
Antenna height	MC: 25 m, SC: 10 m, UE: 1.5 m
Pathloss model	MC: ITU Urban Macro NLOS SC: ITU Urban Micro NLOS
Fading model {MC, SC}	Shadow fading deviation: {6, 4} dB Correlation distance: {40, 10} m
UE speed	6 km/h
Handover model [8]	Event A3, HO margin: 3 dB
Traffic & scheduling	Full-buffer & Resource-fair

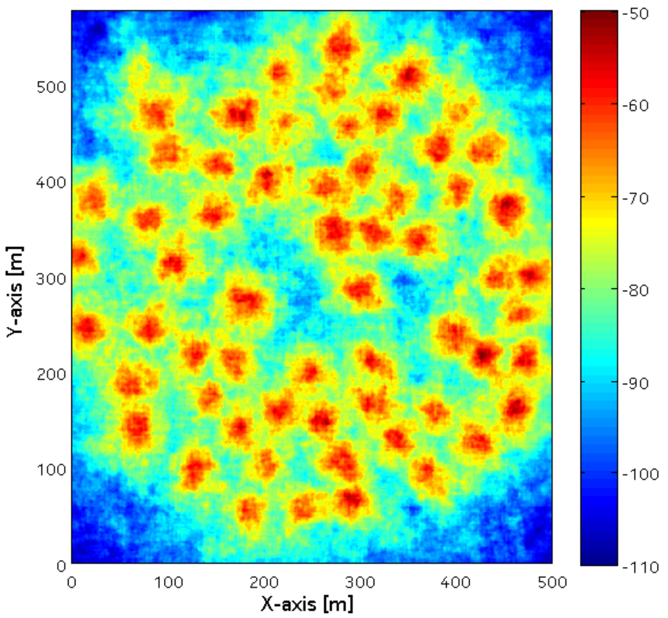

Fig. 2. The coverage of small cell-layer in terms of the best reference signal received power (RSRP) [dBm] when the small cell density is 20 SC/macro.

Signal-to-interference-plus-noise-ratio (SINR) (of UE u connected to cell c) are defined by

$$\gamma_{u,\mathrm{DL}}^{c} = \frac{P^{c} \cdot L_{u}^{c}}{\sum_{k\neq c} P^{k} \cdot L_{u}^{k} + N}, c = X(u), \quad (1)$$

$$\gamma_{u,\mathrm{UL}}^{c} = \frac{P_{u} \cdot L_{u}^{c}}{\sum_{k\neq c}\sum_{i\neq u} P_{i} \cdot L_{i}^{k} + N}, c = X(u), \quad (2)$$

for downlink and uplink respectively. Here, $X(u)$ refers to the connection function of UE u. Cell c can be either an element of small cell-set, $c \in$ SC, or macro-set, $c \in$ MC. Moreover, P^{c} is the transmission power level of serving cell c; P^{k} is the transmission power level of neighboring cell k; P_{u} is the transmission power level of UE u; P_{i} is the transmission power level of interfering UE i; L_{u}^{c} is the signal attenuation between source cell c and UE u; L_{i}^{k} is the signal attenuation between UE i and neighboring cell k; and N is the power level of white noise. It should be noted that the maximal number of UEs multiplexed in a LTE-FDD subframe is limited to 3 in the simulations, whereas the resource-fairness took place in the time domain [7].

Given γ_{u}^{c}, the layer throughput per UE is evaluated for the downlink and uplink respectively by using the Shannon capacity formula because of the fact that there is no specified air interface for 5G and no elaborate modeling thereof,

$$C_{\mathrm{DL}}^{j} = \frac{\sum_{t}^{T}\sum_{c}^{M}\sum_{u}^{N_{c}^{t}} \frac{1}{N_{c}^{t}} \mathrm{BW} \cdot \log_{2}\left(1+\gamma_{u,\mathrm{DL}}^{c,t}\right)}{\sum_{t}^{T}\sum_{c}^{M} N_{c}^{t}}, \quad (3)$$

$$C_{\mathrm{UL}}^{j} = \frac{\sum_{t}^{T}\sum_{c}^{M}\sum_{u}^{N_{c}^{t}} \frac{1}{N_{c}^{t}} \mathrm{BW} \cdot \log_{2}\left(1+\gamma_{u,\mathrm{UL}}^{c,t}\right)}{\sum_{t}^{T}\sum_{c}^{M} N_{c}^{t}}, \quad (4)$$

where BW is the bandwidth; M is the number of cells at layer j; T is the total service time; and N_{c}^{t} is the number of active UEs at cell c at time t.

B. Impact of Network Densification toward 5G

Simulations are performed for three different urban area network densification scenarios toward 2020: 4 SC/macro, 12 SC/macro and 20 SC/macro in order to reflect the trend in performance metrics toward 2020 (e.g., as shown in Figure 3).

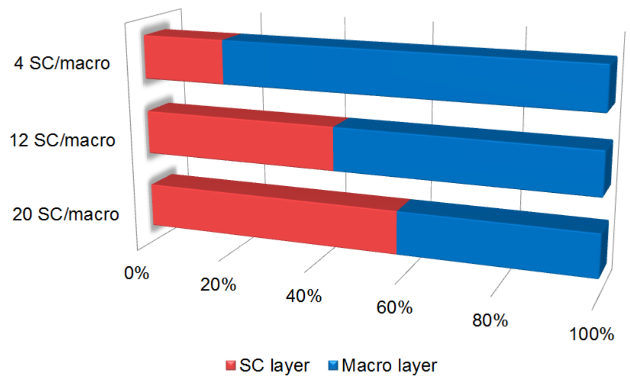

Fig. 3. Offloading rate of macro mobile UEs for three small cell densification scenarios.

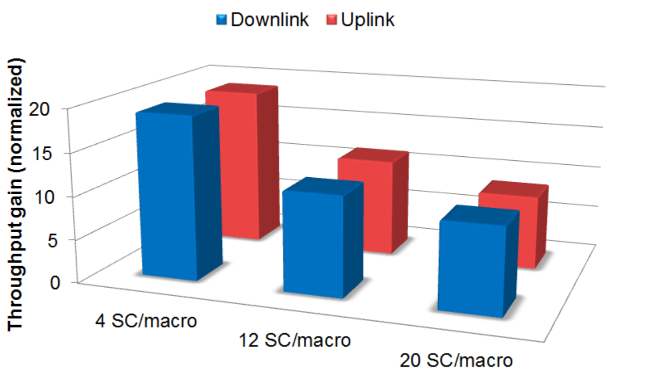

Fig. 4. Normalized throughput gain in downlink ($C_{\mathrm{DL}}^{\mathrm{SC}}/C_{\mathrm{DL}}^{\mathrm{MC}}$) and uplink ($C_{\mathrm{UL}}^{\mathrm{SC}}/C_{\mathrm{UL}}^{\mathrm{MC}}$).

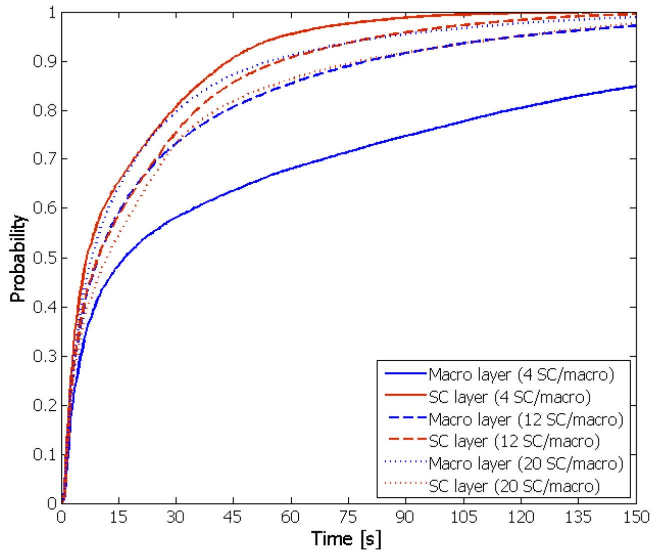

Fig. 5. CDF for the continous connection time per cellular connectivity layer.

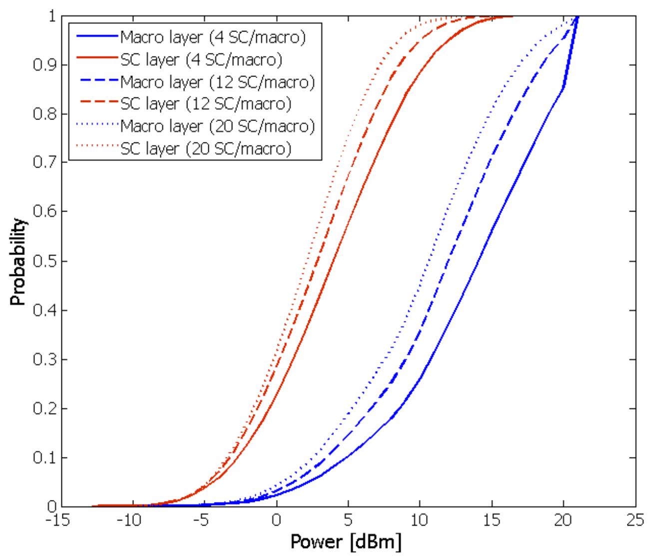

Fig. 6. CDF for the uplink transmission power per cellular connectivity layer.

As seen in Figure 3 on the previous page, it is expected that more than half of the macro mobile data traffic could be offloaded to the small cell-layer at the end of the current decade assuming 20 SC/macro density. Even though the majority of UEs are served by the small cell-layer, small cells are still estimated to be providing notably better throughput than macro cells and having much more available capacity thereof. For instance, at the scenario of 20 SC/macro density, up to 9 and 10 times better mean throughput can be expected for the small cell-layer in uplink and downlink respectively as shown in Figure 4. Hence, even if the small cells were located at hot-spot zones, the network could still benefit from the further offloading from macro cells to small cells, for instance, by scheduling bandwidth-hungry, delay-tolerant cloud services in a larger timescale. On the other hand, it should be noted that the median throughput gain (for the UEs served by small cells) is roughly half of the mean throughput gain based on our analysis.

As it can be observed from the cumulative distribution function (CDF) given in Figure 5, 50%-ile continuous connection time per connectivity layer, i.e., continuous connection time that is not interrupted by an inter-layer handover, is between 6 and 16 s for 6 km/h UE speed. Considering the fact that vehicular communications will play a very important role in 5G [2], the expected continuous connection time will even be much shorter. Thus, as long as the synchronization tolerance is in the order of seconds, bandwidth-hungry, delay-tolerant synchronization tasks can be scheduled to run at the small cell-layer.

While offloading from macro- to small cell-layer, UE benefits from the reduced transmission power which is estimated to be 9 dB in average as shown in Figure 6. It means that the optimized small cell offloading will also help UE significantly preserve battery life. Figure 6 also shows that not only the small cell-UEs but also the macro-UEs take the advantage of lower transmission power as small cells provide larger coverage with larger RSRP (with respect to macro cells) at the network-edge.

III. Mobile Connectivity Management for Cloud Services

A. *Context-Aware Cloud Connectivity Management*

To utilize alternative connectivity options offered by HetNets, besides the channel measurements, mobility-context (e.g., user velocity and proximity to the small cells), traffic-context (e.g., service profile and network load), energy-context

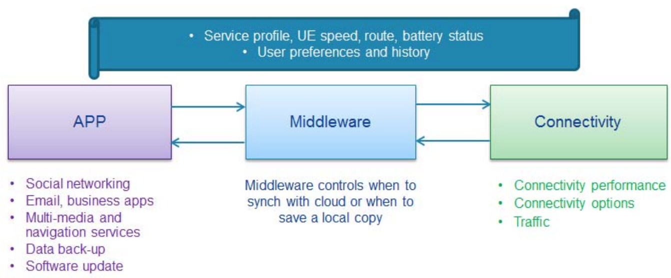

Fig. 7. Information flow between different entities of wireless cloud connectivity: application (APP), middleware and connectivity.

(e.g., battery status) as well as user preferences (e.g., billing) can be taken into account in the mobile connectivity management for bandwidth-hungry, delay-tolerant cloud services. However, this requires the virtualization of the connected device and the cloud services in a heterogeneous connectivity setting as depicted in Figure 7 on the previous page. The major entity of the virtualization, which simplifies the communication between application layer and connectivity layer, is called middleware. Middleware translates the state of one layer to another so that only the required information is provided for the targeted tasks. More specifically, the main role of the middleware is to decide, based on the gathered context information (e.g., mobility and traffic) and policies, whether the synchronization with the cloud should be advanced or delayed.

Context information could be locally obtained (e.g., through sensors) or fetched from the other nodes (e.g., BS and cloud servers). For instance, we propose that the information exchange on the traffic-context takes place between UE and network in both directions. By means of mutually transparent context information, UE can predict the optimal time to schedule a cloud synchronization task. Furthermore, network can reserve dedicated resources in advance and take necessary actions for the traffic steering and load balancing mechanisms. Thus, even very short connections to the small cell layer would be sufficient to manage heavy synchronization tasks with early resource reservation and prioritization capabilities of context-aware networks.

Assuming the use of accurate mobility-context and mutually transparent traffic-context, the offloading probability p_{opt} can be defined by

$$p_{\mathrm{opt}} = P[u \mid X(u) \in \mathrm{SC}, t_o - \Delta \leq t \leq t_o + \Delta]\,, \tag{5}$$

where t_o and t refer to original and alternative synchronization times respectively. Moreover, Δ denotes the synchronization tolerance that depends on the QoS requirements of the cloud service as well as the user preferences. Here, we note that the accuracy of the mobility-context information is crucial when deciding whether the synchronization task should be delayed or advanced. It is due to the fact that an inaccurate estimation may push UE to run synchronization tasks closer to the macro cell-edge and consume more macro-layer resources and larger transmission power thereof.

In the performance evaluation methodology, the normalized UE transmission power consumption L_{TXP} is defined by

$$L_{\mathrm{TXP}} = \frac{p_{\mathrm{opt}} \cdot E[\mathrm{TXP^{SC}}] + (1 - p_{\mathrm{opt}}) \cdot E[\mathrm{TXP^{MC}}]}{p_{\mathrm{ref}} \cdot E[\mathrm{TXP^{SC}}] + (1 - p_{\mathrm{ref}}) \cdot E[\mathrm{TXP^{MC}}]}\,, \tag{6}$$

where $E[\mathrm{TXP^{MC}}]$ and $E[\mathrm{TXP^{SC}}]$ refer to the expected uplink transmission power (for the same amount of data) in macro- and small cell-layers respectively; and p_{ref} denotes the small cell offloading probability when no context information is in use.

B. Performance Evaluations for Context-Aware Cloud Connectivity Management

The simulation methodology to study the introduced concept follows the same simulation modeling principles discussed in Section 2.A. Simulations are performed for 20 SC/macro density assuming three mobility scenarios: *a.* 6 km/h, *b.* 15 km/h and *c.* 30 km/h. In the performance evaluations for the proposed concept, it is assumed that UE mobility, i.e., UE velocity (and route), is correlated during Δ and predictable for the given Δ time interval thereof. However, in practical implementation, there may be errors in mobility prediction for large Δ values. Hence, Δ values investigated in this study are selected rather smaller, i.e., from 1 s to 30 s.

Figure 8 shows the offloading gains for context-aware cloud connectivity management. Based on the results, when Δ is small, significant offloading gains can only be achieved for the users moving faster than pedestrians (e.g., jogging, cycling or traveling). However, if Δ is above 1 s, notable improvements can also be observed for pedestrians.

From the power-consumption point of view, as depicted in Figure 9, it is possible to get significant battery savings in the user terminal especially for delay-tolerant heavy synchronization tasks, such as data back-up. Therefore, as discussed in the previous sections, the battery status can also be a criterion to initiate the context-aware connectivity management schemes, in other words, Δ can be stretched when the user terminal indicates a low battery condition.

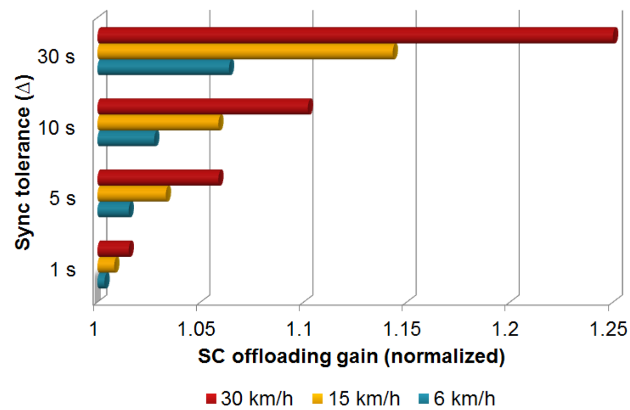

Fig. 8. Offloading gain ($p_{\mathrm{opt}}/p_{\mathrm{ref}}$) when context-aware cloud connectivity management is performed.

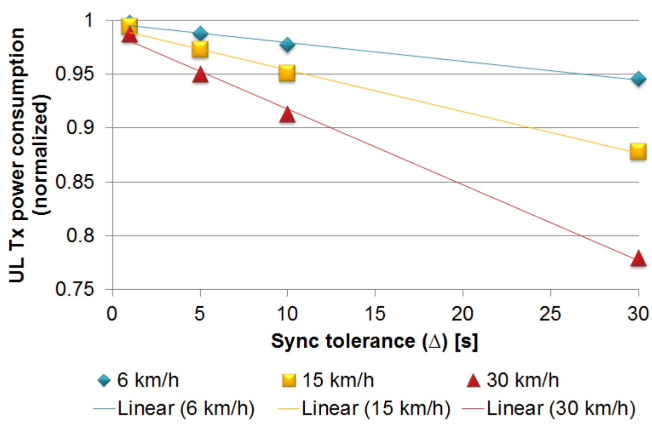

Fig. 9. Uplink transmission power consumption (L_{TXP}) when context-aware cloud connectivity management is performed.

IV. Complementary Cloud Server-Managed Solutions

A. Proactive Resolution Mechanism for Cloud Synchronization Conflicts

When tuning the cloud synchronization time table, potential cloud synchronization conflicts, between multiple devices or users, may occur. In addition, interruption in the broadband connectivity, temporary lack of resources or a parallel service may result in a synchronization conflict as well. To solve this problem proactively, each UE can message to the cloud server about its wireless connectivity characteristics and the estimated synchronization time. As a response, the cloud server may message back about an expected conflict and ask for a re-ordering for the related synchronization tasks (e.g., fast fetching, or postponing). The signaling flow given in Figure 10 illustrates the proposed mechanism, where both t_1 and t_2 are assumed to be the delayed synchronization times within Δ time interval for UE_1 and UE_2 respectively; and $t_2 - t_1 \leq \Delta$.

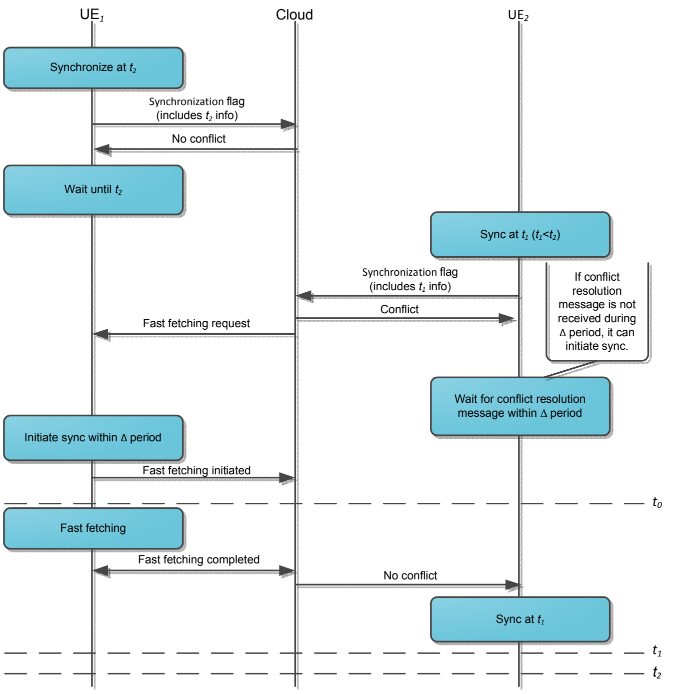

Fig. 10. Signaling flow between multiple UEs (UE_1 and UE_2) and cloud in case of proactive resolution mechanism for cloud synchronization conflicts.

V. Conclusions

Context-aware cloud connectivity management concept introduced in this paper may enable the optimized utilization of alternative connectivity layers when bandwidth-hungry, delay-tolerant cloud services are accessed through the future HetNets. In addition, with the help of cloud server – UE interaction, potential cloud synchronization conflicts can be avoided.

As shown in the simulation results, traffic offloading from macro- to small cell-layer can be further optimized because of the fact that small cells offer macro mobile UEs, which often change connectivity layers in HetNets, better or similar QoS with less power consumption. Moreover, both UE and network can benefit from the more predictable data rate and capacity since they can proactively interact and optimize the mutual data traffic.

Future work may consider more elaborate mobility scenarios and the cloud data traffic modeling in detail.

Acknowledgment

Part of this work has been performed in the framework of the FP7 project ICT-317669 METIS, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues in METIS, although the views expressed are those of the authors and do not necessarily represent the project.

The authors would like to thank Kimmo Valkealahti (Nokia Research Center, Finland) for his assistance in the performance evaluations for the preliminary studies.

References

[1] A. Osseiran, et al., “The foundation of the Mobile and Wireless Communications System for 2020 and beyond Challenges, Enablers and Technology Solutions,” In proceedings of IEEE Vehicular Technology Conference (VTC2013-Spring), Dresden, Germany, June, 2013.

[2] METIS, Deliverable D1.1. Scenarios, requirements and KPIs for 5G mobile and wireless system, May, 2013. Available: https://www.metis2020.com/documents/deliverables

[3] R. Kwan, C. Leung, J. Zhang, “Multiuser scheduling on the downlink of an LTE cellular system,” Research Letters in Communications, vol. 2008, Article ID 323048, 2008.

[4] A. Lobinger, S. Stefanski, T. Jansen, I. Balan, “Load balancing in downlink LTE self-optimizing networks,” In proceedings of IEEE Vehicular Technology Conference (VTC2010-Spring), Taipei, Taiwan, May, 2010.

[5] T. Ihalainen, et. al., "Flexible scalable solutions for dense small cell networks," Wireless World Research Forum (WWRF), Oulu, Finland, April 2013.

[6] 3GPP TR 36.814 V9.0.0, *Further advancements for E-UTRA physical layer aspects (Release 9)*. Available: http://www.3gpp.org/ftp/Specs/html-info/36814.htm

[7] 3GPP TS 36.213 V11.4.0, *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 11)*. Available: http://www.3gpp.org/ftp/Specs/html-info/36213.htm

[8] 3GPP TS 36.331 V.11.5.0, *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (Release 11)*. Available: http://www.3gpp.org/ftp/Specs/html-info/36331.htm