

An Architecture for Energy-aware On-demand Mobile Network Management

Manuel Peuster
University of Paderborn
manuel.peuster@upb.de

Holger Karl
University of Paderborn
holger.karl@upb.de

ABSTRACT

The increasing amount of mobile traffic leads to a significantly higher energy consumption of mobile networks that is mainly caused by the high number of required base stations. One recent solution for this is based on a two-layered network that uses long-range macro cells to provide a full coverage signaling overlay and short-range small cells for fast data transmissions. These small cells can be switched off when they are not needed and allow network-wide energy optimizations.

This paper presents an architecture that extends existing mobile networks to integrate a small cell layer that supports on-demand cell activation. We discuss how additional small cells can be interconnected with existing core components and how they can be controlled by a resource management component. Finally, a Wi-Fi based proof of concept testbed implementation is presented that demonstrates the feasibility of the approach.

CCS Concepts

•Networks → Network management; *Network manageability*;

Keywords

Cell-switching; HetNet; MPTCP; small cells; testbed

1. INTRODUCTION

The amount of Internet traffic that is delivered by mobile networks has dramatically increased during the last couple of years and this trend is predicted to continue. Compared to 2006 the overall data traffic in mobile networks has grown by about 400% [4] and more than 15

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AllThingsCellular'15, August 17-21 2015, London, United Kingdom

© 2015 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-3538-6/15/08...\$15.00

DOI: <http://dx.doi.org/10.1145/2785971.2785973>

exabytes of mobile traffic will be delivered per month in 2018 [5]. This forces mobile operators to increase the capacity of their networks to satisfy the needs of their subscribers with the downside that the amount of energy consumed by mobile networks is heavily increasing.

Consequently, a lot of research effort was spent to develop new technologies for future networks. One solution is deploying networks that consist of different kinds of base stations, so called heterogeneous mobile networks (HetNets). These networks are built of conventional macro cells and additional small cells with limited transmission power that are deployed with much smaller inter-site distances. These small cells are optimized for short-range, high-rate transmissions and provide the needed capacities to the users with relatively low energy consumption. Nevertheless, further energy saving techniques are needed to reduce the overall energy consumption of future networks. One recent idea splits mobile networks into a signaling and a data layer to which the user equipment (UE) is connected simultaneously. The signaling layer is formed by high-range low-rate macro cells and the data layer by low-range high-rate small cells. This approach creates the possibility to switch off unneeded small cells while maintaining full service coverage with the signaling layer [3].

This paper proposes a way to integrate such a two-layer architecture with separated responsibilities into an existing mobile network deployment based on the evolved packet system (EPS) [1]. The macro cell layer in our approach is used for signaling as well as for slow user data transmissions and the small cell layer is only used for user traffic. The presented approach extends the existing core and integrates additional components that allow network-wide resource optimizations. We discuss this integration down to the level of interaction patterns which goes beyond existing architecture proposals. The general feasibility of this design is shown with a proof of concept implementation that emulates small cells with Wi-Fi access points.

2. RELATED WORK

The idea of using small cells to create a separated network layer that is only used for user data transmissions was recently proposed in several publications. The

main focus of these two-layered networks is improving the overall network capacity by optimizing macro cells for long-range low-rate transmissions of signaling data and small cells for short-range high-rate transmissions of user data. But it is also a key enabler for energy-efficient mobile networks and investigated by green research projects, like GreenTouch [8].

Capone et al. [2, 4, 12] utilize this two-layer approach to describe an energy efficient network design that uses a network-wide optimization approach to save energy by switching off small cells. Their approach uses extended context information which has to be provided to the network and an additional resource management entity which is needed to optimize the network resource usage and assign UEs to appropriate small cells. They do not provide any technical solutions nor details about the integration of their proposed architecture into an existing mobile network.

A more detailed architecture description for a network that separates its signaling and data layer is given in [9]. They introduce the so-called phantom cell concept with small cells which do not send any pilot signals. These cells are directly connected to a macro cell and form one logical base station from the perspective of a UE. The authors of [13] extend this phantom cell approach and focus on energy savings achieved by switching off small cells. They introduce a database to store past signal measurements between UEs and small cells, providing the possibility to determine which cell is most suitable to be switched on for a given UE. Nevertheless, they do not consider network wide optimizations, like done in our approach.

The authors of [6] use a central component for assignment decisions to a data layer created with Wi-Fi hotspots. Their decisions are based on extended state information collected by an application that is executed on each UE. Nevertheless, their optimization goal is an improved user experience and they do not consider energy efficiency by switching off parts of their network.

All presented approaches use a separation between signaling and data traffic. But the idea of switching off small cells with a central decision entity and thus optimize the networks's energy consumption on system level is only considered by [2, 6]. This paper extends these ideas and proposes a complete architecture that integrates a separated small cell layer as well as a resource management system into an existing EPS-based network. We use an interconnection pattern between core and small cells which is close to the phantom cell approach, with the difference that small cells are visible to the UEs when they are switched on and do not form one logical cell together with a macro cell.

3. ARCHITECTURE

Our architecture proposal starts with the integration of additional small cells and resource management components. After this, we discuss a novel intercon-

nection pattern that introduces resource management planes on existing interfaces. At the end, procedures are presented which handle the activation, assignment and handover events in our two-layered network design.

3.1 Network Components

At first, additional small cells have to be integrated into an existing network in order to create a second layer for fast data transmissions. One solution for this are remote radio heads (RRH) which are connected to a macro base station that performs the baseband processing. This would allow the tight integration of the data layer and fine-grained control over its base stations. However, there are two problems with this approach. First, fast fiber connections with low latencies are needed between each RRH and macro base station. Second, the number of RRHs that can be served by a single macro base station is limited due to the available compute power needed for the baseband processing at each base station site. As a result, data layer small cells should be designed as standalone basestations which can be connected to the core by already available connections, such as copper or microwave links. These connections will then be used for user data transmissions and for control signals to manage the power state of each individual cell. A candidate interface for this is the X2 interface that offers a control and user data plane. Additionally, direct S1-U connections for user data should be established between small cells and core.

As a second step, a resource management component has to be integrated into the core that is responsible for small cell power management and decisions about the assignment between UEs and available small cells. One option for this is integrating this component into an already existing EPS component, like the MME. However, this would cause changes in existing implementations and would imply that for each MME an additional resource manager is added to the core. A better solution is adding only few resource manager components to the network, each responsible for the resource usage optimization in a large part of the network. An example for this could be one resource manager that optimizes the small cell usage of a complete city.

Fig. 1 shows the described deployment consisting of a macro base station (classical eNodeB) that controls a subset of small cells. This is done with a small cell manager (SCM) that is added to each macro cell and is responsible to control the power state of small cells and to manage the communication between resource manager (RM) and UEs. The RM forwards its power management decisions to the SCMs which are then in charge to execute them. These SCMs can additionally be used to aggregate context updates of the UEs. This reduces the number of requests that have to be handled by a single RM and thus improves the scalability. Another important aspect of this design is that the only point of contact between existing EPS components and new resource management components is between the SCM

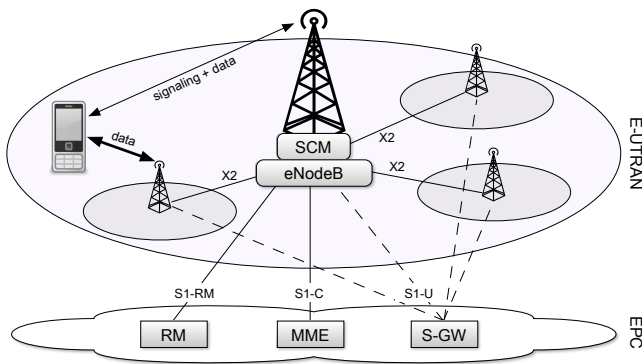


Figure 1: EPS based architecture overview with extended E-UTRAN and EPC

and the eNodeB. An SCM can be implemented as software extension without requiring hardware changes or additional physical connections. This allows an easy and cheap integration of the proposed approach into already deployed networks. It is also possible to deploy the second layer only in some areas of the network, e.g. in areas with high user densities, because existing base station can still operate in single layer mode. This architecture still allows the connection of older UE generations that do not support dual connectivity because the existing EPS architecture is not changed but only extended and macro base stations are still able to transmit data with low rates. This means that older UEs can still interact with macro base stations, like in the unchanged version of the network, but without any advantages provided by the small cell layer.

3.2 Interfaces

There are two additional areas of communication in the proposed EPS extension. The first one is the communication between the RM, SCM, and small cell that is needed to control the small cells and switch them on and off. The second area consists of extended status updates that are sent by UEs to the RM.

The power management of small cells is controlled by the RM that decides which small cells are switched on and off. The decision is then transferred to the corresponding SCM which controls a subset of small cells. This is done by S1 links for the communication between RM and SCM as well as X2 links for the communication between SCM and small cells. There are two design alternatives to implement this. The first one is extending existing EPS application layer protocols, resulting in a solution where the RM pushes its results to an MME which transfers it over an extended S1-AP protocol to the eNodeBs and thus to the SCMs. The SCMs can then use an extended version of the X2-AP protocol to send switching commands to small cells. The second, and preferred, alternative is to use separated control protocols on top of existing interfaces. This creates the possibility for direct communication between the

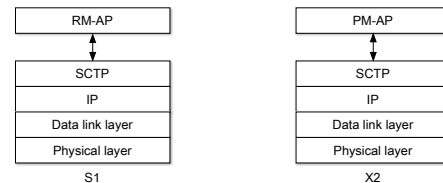


Figure 2: Additional protocol stacks

RM and the SCMs without the need of an intermediate MME which terminates the existing S1-AP protocol at the core. It also increases flexibility and makes the integration into existing deployments cheaper because no updates of existing EPS components are needed in order to handle new versions of the S1-AP or X2-AP protocol.

Fig. 2 shows the protocol stacks of this solution. For RM to SCM communication the Resource Management Application Protocol (RM-AP) is added which runs on top of an additional SCTP connection established between each SCM and the central RM (S1-RM in Fig. 1). This connection can easily be established because of the all-IP design of the EPS that interconnects all components by an IP-based network. The same is done on the X2 interfaces between eNodeBs and small cells where the Power Management Application Protocol (PM-AP) is added that is transported by an additional SCTP connection. This forms a third plane which operates in parallel to the existing control and user plane on the X2 interface.

3.3 Procedures

The proposed architecture moves the decision on how to assign UEs to small cells to the resource manager and changes some default interaction patterns between UE and network. Especially attachment and handover procedures that include the additional small cell layer are needed. We assume that each UE periodically sends status reports to the network so that the resource management component has always enough information to make optimization decisions e.g. to assign a UE to a small cell to fulfill a data request.

3.3.1 Activation and Attachment

The message sequence chart in Fig. 3 shows how the RM activates a small cell and assigns a UE to it. First, the UE sends its status reports to the eNodeB to which it is currently connected and thus to the SCM that is integrated in this eNodeB. The SCM forwards the status information to the RM that runs an optimization algorithm. After solving the optimization problem, the RM returns the *activation and assignment profile* to all involved SCMs. This information is used by each SCM to trigger the connection setup between a UE and a small cell.

There are two cases that must be handled by a SCM when a UE is assigned to the small cell layer. The first

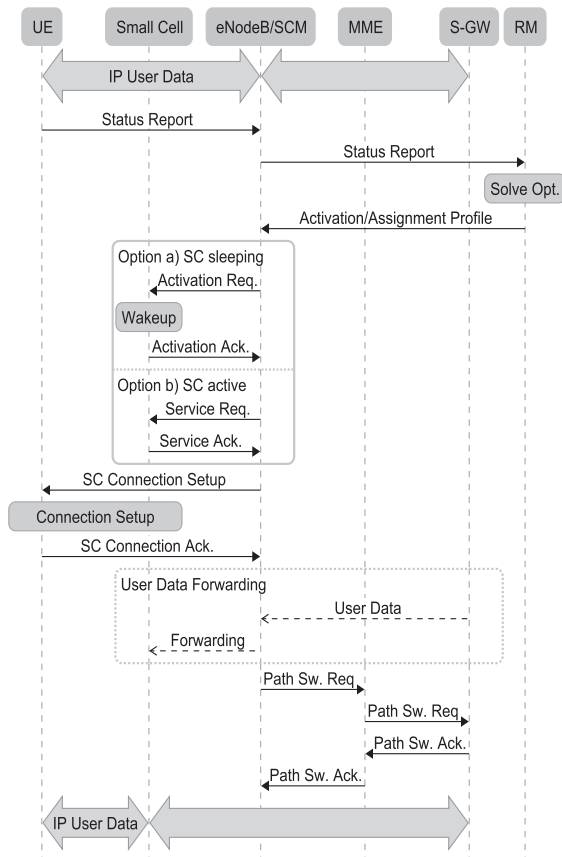


Figure 3: Small cell activation and assignment procedure

case wakes up a sleeping small cell by sending an *activation request* which is acknowledged after the small cell is finally running. The second case deals with an already running small cell which is informed about an upcoming connection by a *service request* and acknowledges it to the SCM. Now, the small cell is ready to be used and the SCM informs the UE to connect to this particular small cell by sending a *small cell connection setup* message. The UE creates an additional connection to the small cell and sends a *small cell connection acknowledgement* message to the SCM. After this, all user data for the UE is forwarded from the source eNodeB to the target small cell via the X2 interface and all user data is delivered by the small cell. As last step, the eNodeB/SCM triggers a *path switch request* at the MME in order to move the user data tunnel from the eNodeB to the small cell. This command is forwarded to the S-GW by a *modify bearer request*. At the end, the MME acknowledges the path switch to the eNodeB and user data is directly transferred from the S-GW to the small cell and from there to the UE.

3.3.2 Handover

Mobility management and thus handovers between cells is a key feature of mobile networks. This task

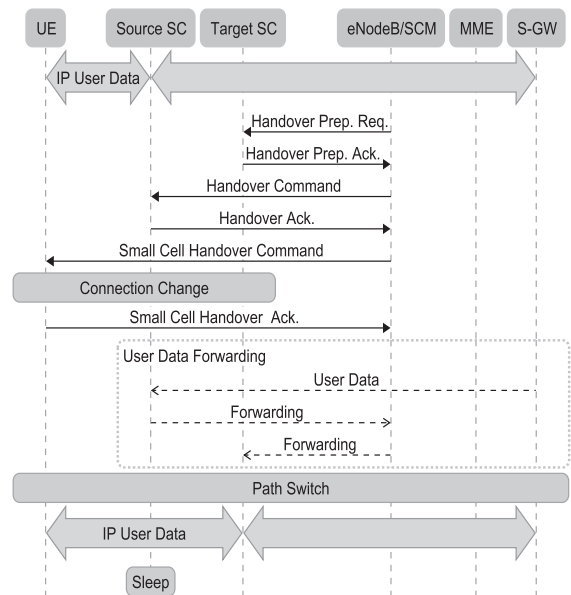


Figure 4: SC-to-SC handover procedure

must also be handled in the proposed two-layer network architecture where especially handovers between small cells happen very often because of their limited ranges. There are three possible types of handover scenarios which can happen in a two-layer network. The first case covers the scenario of a UE which is only connected to the signaling layer and thus only moves between macro cells (eNodeBs). This type of handover can be handled by using the default handover procedures of the EPS and is denoted as MC-to-MC handover. It does not involve any small cells and no changes or further procedures are needed. The second scenario is denoted as SC-to-SC handover which happens when a user moves between small cells without leaving the current macro cell. The third scenario is the most complex one and describes a combination of SC-to-SC and MC-to-MC handover. In this scenario the user moves to a target small cell which is controlled by another macro cell and thus the signaling connection of the UE must also be moved to the new cell.

A solution for the SC-to-SC handover is shown in Figure 4. It starts after the RM has decided to move the UE to another small cell and has informed the SCM about this change. Because of this, a *handover preparation request* is sent from the SCM to the target small cell in order to prepare it for an upcoming connection request from a UE. After this, the SCM sends a *handover command* to the source small cell which triggers data forwarding to the target small cell. This forwarding can be done in two ways. The first one is shown in the figure and forwards user data from the source small cell to the eNodeB/SCM which again forwards it to the target small cell by using the X2 interfaces which are mandatory between each small cell and

eNodeB/SCM. However, implementing this forwarding directly between source and target small cell is also possible if there exists a direct X2 connection between them (compare S1 and X2 handovers in the EPS standard). When the user data forwarding is active, the eNodeB/SCM sends a *small cell handover command* to the UE triggering the connection change to the target small cell. After this, the UE is served by target small cell and the eNodeB/SCM can start to reconfigure the user data tunnel by sending a *path switch request*.

The third scenario is the handover between two small cells not controlled by the same eNodeB/SCM that can be divided into two phases. At first, a handover between small cells is executed that uses forwarding between the involved eNodeB/SCMs. Second, a classical handover from source to target eNodeB is performed in the signaling layer. The downside of this approach is the mandatory X2 connection between all involved eNodeBs that is used to forward the additional signaling messages.

4. TESTBED IMPLEMENTATION

We developed a testbed to show the feasibility of the proposed architecture. The testbed uses Wi-Fi access points to emulate small cells and utilizes a publicly available cellular connection as full coverage signaling layer. This results in two independent network layers without the need to deploy real small cell hardware. This setup does not implement a full EPS core but it provides the minimal necessary environment for experiments with management strategies in a two-layer network. From the perspective of our resource manager, a signaling layer is available over which status updates are received and a couple of small cells can be activated for fast data transmissions. From the perspective of user applications, both network layers are hidden behind a single TCP connection which would also be the case in a real EPS based two-layer deployment. This is achieved by using MPTCP [11] as an instrument to utilize both interfaces for one connection.

4.1 Testbed Setup

Fig. 5 shows the testbed setup. It shows a user who carries a smartphone and a location tracking device; both communicate with a backend that is responsible to control the whole system. An additional management service is running on the smartphone and collects status information of the device which is sent to the backend by using the cellular connection.

The backend component decides which of the access points is activated and how the user devices are assigned to them. This decision is taken by a resource management plugin that runs inside the controller and can access all status information provided by the UEs. This includes the exact location of each user which can be replaced by signal measurements in real-world deployments. Since GPS-based location data is not available

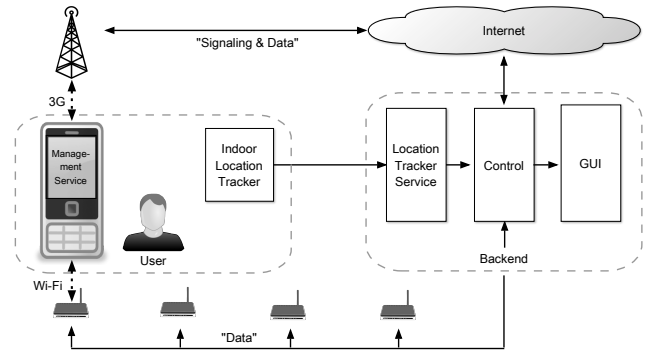


Figure 5: Testbed setup

when our testbed is placed indoors, a third-party location service, like LAURA [10], can be used. The backend’s API interface is available from the Internet so that it can be accessed by the UEs via their cellular connection. For demonstrations, a web-based GUI is available that visualizes the system state.

As UEs, Samsung Galaxy Nexus I9250 smartphones are used which run a rooted version of Android. Each device runs a custom management service that activates and manages the Wi-Fi interface and cellular interface simultaneously. This is not possible with the stock version of Android. The management component also controls the routing setup so that it ensures that all signaling traffic that is sent to the backend is always transmitted via the cellular interface. In addition, the complete system state of the device is monitored by the management service and is periodically transmitted to the backend. These state updates include current network traffic, display state, and active application and may be used by a resource management plugin in order to determine which user devices are active and should be connected to the Wi-Fi data layer e.g. when they exceed a certain traffic threshold. We ported the original MPTCP kernel implementation to the Galaxy Nexus device in order to use both network layers for one data transmission¹.

4.2 Results

We demonstrate the functionality of the testbed with a simple experiment that shows the on-demand activation, assignment changes, and deactivation of access points. We use a resource management plugin that enables the closest access point when it detects a data transmission on the mobile interface of a UE. Fig. 6 shows the throughput of the cellular and Wi-Fi interface during the experiment in which a file is downloaded from an MPTCP enabled server. During the file download, the location of the UE was changed twice so that the backend reassigns it to another access point which can be seen around second 110 and 205 where the Wi-Fi throughput drops for a short time when the connection

¹<https://github.com/mpeuster/mptcp-galaxy-nexus>

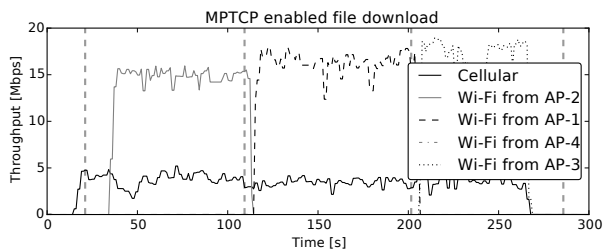


Figure 6: Cellular and Wi-Fi throughput of the mobile device during 500 MB file download with enabled MPTCP

is moved between access points. The decisions taken by the backend are marked with vertical, dashed lines. The first marker shows the decision of the backend to use Wi-Fi as additional connection. After this, the first access point is activated and the UE is served by the mobile and Wi-Fi network simultaneously by using a MPTCP connection with two subflows. The second and third marker show the point in time when the backend receives location changes of the UE and reassigns it to other access points. The last marker shows how the backend deactivates the last active access point.

The experiment shows how the testbed can be used to test the behavior of a simple threshold based resource management strategy. It is not thought as a detailed performance or energy efficiency analysis which is left for future work. Detailed simulation results on potential energy savings based on small cell switching are presented by the GreenTouch mobile working group [7].

5. CONCLUSION

This paper proposes an architecture that shows how existing mobile networks can be extended in order to support on-demand small cell activation with separate signaling and data layers. Our approach adds a number of components to the EPS and discusses how they can interact with existing parts of the core. Most of our changes are software-based and use standardized interfaces so that already deployed networks can easily be updated. We added a central resource manager that controls assignment and activation decisions to optimize the network, e.g. for energy efficiency. Existing management components, like MMEs, are not changed so that the proposed architecture is backward compatible and supports today's UEs which are not able to connect to both layers at the same time. It is also possible to deploy this architecture only in some areas of the network and continue single-layer operation in other areas. This makes the proposed approach very flexible and reduces deployment costs. A testbed implementation is presented that can be used to test different resource management strategies in a realistic usage scenario. The testbed can be extended for further experiments on two-layer network designs as well as data offloading approaches that utilize Wi-Fi hotspots.

6. ACKNOWLEDGMENTS

This work was partially supported by the GreenTouch project [8] and by the German Research Foundation (DFG) within the Collaborative Research Centre "On-The-Fly Computing" (SFB 901). The testbed was developed in cooperation with the GreenTouch BCG2 demo project partners from the Politecnico di Milano and University of Piraeus Research Centre.

7. REFERENCES

- [1] 3GPP. Technical Specification Group Services and System Aspects; Network architecture. TS 23.002, 3rd Generation Partnership Project (3GPP), 2014.
- [2] A. Capone, I. Filippini, B. Gloss, and U. Barth. Rethinking cellular system architecture for breaking current energy efficiency limits. In *Sustainable Internet and ICT for Sustainability (SustainIT), 2012*, pages 1–5. IEEE, 2012.
- [3] A. Capone, A. Fonseca dos Santos, I. Filippini, and B. Gloss. Looking beyond green cellular networks. In *Wireless On-demand Network Systems and Services (WONS), 2012 9th Annual Conference on*, pages 127–130. IEEE, 2012.
- [4] Y. Chen, O. Blume, A. Gati, A. Capone, C.-E. Wu, U. Barth, T. Marzetta, H. Zhang, and S. Xu. Energy saving: Scaling network energy efficiency faster than traffic growth. In *Wireless Communications and Networking Conference Workshops (WCNCW), 2013 IEEE*, pages 12–17, 2013.
- [5] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018. 2013.
- [6] A. Dogadaev, A. Checko, A. P. Avramova, A. Zakrzewska, Y. Yan, S. Ruepp, M. Berger, L. Dittmann, and H. Christiansen. Traffic Steering Framework for Mobile-Assisted Resource Management in Heterogeneous Networks. In *ICWMC 2013, The Ninth International Conference on Wireless and Mobile Communications*, pages 107–112, 2013.
- [7] GreenTouch Mobile Communications WG. White Paper: GreenTouch Technical Solutions for Energy Efficient Mobile Networks. 2015.
- [8] GreenTouch Project. <http://www.greentouch.org/>, 2014.
- [9] H. Ishii, Y. Kishiyama, and H. Takahashi. A novel architecture for LTE-B: C-plane/U-plane split and Phantom Cell concept. In *Globecom Workshops (GC Wkshps), 2012 IEEE*, pages 624–630. IEEE, 2012.
- [10] LAURA: Localization and Ubiquitous Monitoring of Patients for Health Care Support. <http://laura.como.polimi.it/>, 2014.
- [11] C. Raiciu, C. Paasch, S. Barre, A. Ford, M. Honda, F. Duchene, O. Bonaventure, M. Handley, et al. How Hard Can It Be? Designing and Implementing a Deployable Multipath TCP. In *NSDI*, volume 12, pages 29–29, 2012.
- [12] A. Redondi, I. Filippini, and A. Capone. Context Management in Energy-efficient Radio Access Networks. In *Digital Communications-Green ICT (TIWDC), 2013 24th Tyrrhenian International Workshop on*, pages 1–5. IEEE, 2013.
- [13] E. Ternon, P. Agyapong, L. Hu, and A. Dekorsy. Database-aided Energy Savings in Next Generation Dual Connectivity Heterogeneous Networks. *IEEE WCNC*, 2014.