Computer Communications 109 (2017) 184–201

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

A context extraction and profiling engine for 5G network resource mapping

Panagis Magdalinos [a,*], Sokratis Barmpounakis [a], Panagiotis Spapis [b], Alexandros Kaloxylos [b], Georgios Kyprianidis [a], Apostolis Kousaridas [b], Nancy Alonistioti [a], Chan Zhou [b]

[a] University of Athens, Athens, Greece
[b] Huawei Technologies, Munich, Germany

ARTICLE INFO

Article history:
Received 21 April 2016
Revised 11 January 2017
Accepted 12 June 2017
Available online 16 June 2017

Keywords:
5G network
User profiling
Data mining
Network optimization

ABSTRACT

Future 5G network ecosystems comprise a plethora of 3GPP and non 3GGP Radio Access Technologies - RATs. Deployment scenarios envision a multi-layer use of macro, micro and femto-cells where multi-mode end devices, supporting different applications, are served by different technologies. The association of end devices to the most appropriate RAT/layer will therefore become a tantalizing process necessitating the introduction of mechanisms that decide and execute an optimal mapping. The latter is of paramount importance since sub-optimal configuration of network components will affect overall network performance. Towards this end, we introduce the Context Extraction and Profiling Engine (CEPE), a knowledge discovery (KDD) framework catering for the extraction and exploitation of user behavioral patterns from network and service information. An eNB exploits the knowledge scheme derived by CEPE in order to improve the placement of end devices to RATs/layers. In the context of this paper, we provide a thorough analysis of existing standards, research papers and patents, discuss the main innovation of our proposal and highlight the differences with existing schemes. Building on use cases involving mobility management mechanisms that typically affect device to technology mapping (i.e. cell (re)selection, handover) we provide an extensive set of experiments that demonstrate the validity and viability of our idea. Overall evaluation showcases that CEPE achieves high quality results thus emerging as a viable approach for network optimization in future 5G environments.

1. Introduction

Future 5G networks target to support a plethora of 3GPP (GSM, HSPA, LTE, LTE-A) and non 3GGP Radio Access Technologies - RATs - (e.g., WiFi). According to network traffic data analysis and projections [1], the envisaged deployment scenarios consist of a multi-layer use of macro, micro and femto-cells where multi-mode end devices, supporting different applications, are served by different technologies. Besides the tremendous growth, which is expected in terms of number of devices, due to an increasingly diverse set of new and yet unforeseen services, users and applications (including machine-to-machine modules, smart cities, industrial automation, etc.), novel and less predictable mobile traffic patterns are also expected to emerge [2]. In such context, the inherently complex task of associating end devices to the most appropriate RAT/layer necessitates special care, since in this ultra dense environment any sub-optimal configuration of network elements will affect overall network performance. Note that although the fundamental problem of placing UEs to the most appropriate RAT is a well-studied issue for past cellular generations, it is expected to be of paramount importance specifically for 5G networks where the expected much higher densification of the network will offer multiple points of attachments of the UEs to the network.

In cellular networks, three mechanisms affect the appropriate placement of end devices to RATs/layers, namely: cell (re) selection, call admission control and handover. Cell (re)selection is a device control operation while call admission control and handover (in the case of horizontal handovers) are network controlled operations assisted by end devices.

Due to their importance, these mechanisms have drawn significant attention from the research community. During the past years, standardization bodies like 3GPP have specified well established procedures that typically employ simple algorithms (e.g., an end device or an eNB evaluates the signal strength) to reach a decision. Researchers worldwide have built on these solutions and produced a rich set of algorithms both in scientific publications and patents. In order to provide more sophisticated mechanisms the proposed

* Corresponding author.
E-mail address: panagis@di.uoa.gr (P. Magdalinos).

http://dx.doi.org/10.1016/j.comcom.2017.06.003
0140-3664/

approaches take into consideration additional parameters like the location and speed of a terminal, the experienced interference, the executed service, the required Quality of Service (QoS), the available bandwidth, the energy consumption, the user profile etc. All this information is referred as "contextual information" and can be used to improve network performance and eventually the QoE for the users. The main goal of all these mechanisms is to employ appropriate tools (e.g., utility functions, fuzzy logic etc) so as to evaluate the context information and reach a decision which optimizes the placement of the users in RATs and layers in terms of throughput/latency/delay or other KPIs.

An alternative approach is to monitor the behavior of a user (e.g., location, mobility pattern, use of specific services), analyze it and try to create a classification of end devices based on the respective user behavior. This classification can then be exploited by the mechanisms that affect the placement of end devices to RATs/layers. For example, there are users that use their smartphones only for placing phone calls, while others are "demanding" data users (e.g., web surfing, emails, games). For the first case, the network could place devices on legacy systems like for example GSM while for the latter they could be placed on LTE or WiFi access networks.

The innovation of this idea is to build a user profile on an automated way by analyzing in offline mode a number of user-related parameters and combine them with available contextual information. In order to achieve this, knowledge discovery schemes (i.e. data mining and machine learning techniques) are suitable enablers to collect vast amount of information and from them automatically extract the expected behavior of end users. Note that although the bibliography has a plethora of solutions on placing UEs to RATs, usually the generation of a user profile in an automated way is not discussed and the alternative solutions assume that this information is somehow collected. The main contribution of our paper is exactly the design and evaluation of a knowledge discovery framework capable of extracting such information and using it in network control functions to improve the overall performance of the network.

In the context of this paper we provide a thorough analysis of existing mechanisms (i.e., cell (re)selection, call admission control and handover) as they are specified in 3GPP standards, research papers and patents (Section 2). Following the analysis, we discuss how these approaches differ from our proposal of building a user profile from contextual information and using it to improve control mechanisms in cellular networks. We also detail the procedure of extracting knowledge from context information and present our approach for building a Context Extraction and Profiling Engine (CEPE) and using it (Section 3). We assess the validity and viability of the proposed mechanism by means of extensive experimentations under real world scenarios (Section 4). We conclude the paper and sketch future research directions in Section 5.

2. Related work

In the context of this paragraph we present existing mechanisms for cell (re)selection, call admission control and handover as specified in 3GPP standards, research papers and patents (paragraph A). We also provide a small overview of the various KDD tools that we use later on in CEPE (paragraph B). We conclude the section and pave the way towards the definition of CEPE by providing a motivating discussion in paragraph C.

2.1. RAT/Layer mapping mechanisms

2.1.1. Cell selection/reselection

The cell selection procedure, as specified in 3GPP [3], is based on the link quality level indicators, namely the Reference Signal Received Power (RSRP) and the Reference Signal Received Quality (RSRQ). Several approaches have been proposed in the literature, primarily targeting a) reduction of number of handovers, b) minimization of the admission control reject probability upon a new session initialization as well as c) handover failures. All methods [4–19] consider signal strength-related information (RSS, RSRP, RSRQ, SINR, etc.) while in most of them, a bias in favor of the small cells is used. Almost all approaches are decentralized, apart from [7], which exploits the User Equipment's (UE) RSRP measurements and proceeds in power control for interference management.

A considerable number of efforts attempts to optimize the derived solution based on interference mitigation [8–11] or other power control techniques [7,8,11,12]. ABS schemes (Almost Blank Subframes- employed for reducing interference) are often used [13–15], in some cases for the identification of the available/expected throughput. The expected bit-rate is another input, which several proposals are considering so as to provide the required user requirements [4,12–17]. In parallel to the academic research effort, several patents have been claimed in relation to cell (re)selection schemes [20–24]. Similarly to the aforementioned solutions proposed in literature, most patent focus in simple RSS/RSRP-based solutions, while all propose decentralized mechanisms, i.e., mechanisms, which are deployed on the UE side, where the primary processing and decision making takes place; the UE selects the optimal target RAT/cell and makes the respective request towards the base station and/or the network core. However, the patent proposals are in general simpler and easier to implement since they are mainly based on smaller number of parameters and the information exchange involves less network elements.

In general, context information related to the UE behavior (e.g., UE mobility, accessed services, etc.) is rarely employed for cell selection. The schemes that use context information focus only on the potential target cell's available bandwidth, or the probability of being served by a specific cell. To the best of our knowledge, there is no method that exploits detailed past information of UEs for cell (re)selection.

2.1.2. Call Admission Control (CAC)

According to 3GPP, call admission control procedure involves the RRC connection establishment procedure, which is triggered by either the UE or the network. The topic of call admission control is one of the best studied in the literature. The survey in [25] proposes a classification of user-based call admission control policies, while it identifies some significant features towards the CAC procedure optimization such as channel utilization maximization and QoS reduction minimization. In another survey [26], the authors attempt to make a new categorization of the proposed solutions: prioritized, non-prioritized and optimal policies. Different categorizations of proposed approaches have also been made, such as in [27], where deterministic/stochastic guarantee aspects, distributed/local control and adaptation to traffic conditions are taken into account. Additionally, several CAC schemes are compared in terms of performance and complexity; the common characteristic of all these schemes is the handover prioritization.

Besides these surveys, several additional schemes [28–37] attempt to address CAC related challenges. Most of them operate at the eNodeB/Wi-Fi Access Point level [28,30,32–35], while the rest provide a general overview of the network. Regarding context information, most mechanisms take into account available bandwidth, while fewer consider the type of the service [34,35], and UE mobility [37]. Concerning patents [38–43], most claimed schemes perform CAC-based on the base stations local view (bandwidth, backhaul link, location, etc.). We also observed that several solutions assume QoS classes [44–46]. Finally, only two mechanisms [39,47] are exploiting previous knowledge, which however is used only for bandwidth estimations without taking into account the generic user behavior.

2.1.3. Handover

A handover is the process of transferring an ongoing call or data session from one channel connected to the core network to another channel. A handover can take place from an LTE eNB to another eNB, to a femtocell (HeNB), or to a completely different Radio Access Technology (RAT) within 3GPP standards (e.g. UTRAN technology network), or even out of the 3GPP specified technologies (e.g., Wi-Fi). The second type of handover is characterized as vertical handover.

The primary handover policy within 3GPP is the "Strongest cell handover decision policy", based on the RSRP or RSRQ values measured by the UE. Small cells are considered as one of the most promising solutions for macro-cellular network layouts. Towards this end, 3GPP's LIPA and SIPTO standards [48] aim to improve the femto-macro cells' interworking and offloading. Numerous handover mechanisms have been proposed, while much effort has been invested into further optimizing the standardized mechanisms. Xenakis et al. present [49] an overview of the main criteria and propose a classification which shows that a significant number of efforts emphasize on the received power (RSSI, RSRP, RSRQ, etc.), without taking into account additional contextual information such as the mobility of the UE, or the type of the UE's traffic.

On the other hand, researchers have proposed multiple solutions [50–58], some of which target the mitigation of interference, the spectral efficiency of the user or related issues [50,51,53,54], others make use of SDN-controlled access networks [52]; other use diverse context parameters in order to build a holistic user profile [55–58]. In another comprehensive survey [59], the authors present a plethora of vertical handover mechanisms based on several input parameters, such as latency, RSS, SINR, packet loss, throughput, bandwidth, number of connected users, preferred network, mobility, as well as battery consumption. Many of these mechanisms attempt to create an overall context-aware mechanism, by combining several of the afore-mentioned parameters for the vertical HO decision outcome.

In addition to the schemes proposed in literature, a significant number of patents related to handover optimization mechanisms have been proposed [60–66]. Some emphasize on context-related information collection, while others focus on several other technical enhancements. An initial observation is that some proposals attempt to enhance the handover procedure's core procedures; either from the algorithmic or the technical perspective, i.e., by modifying the triggering mechanisms, the RSRQ threshold schemes, the exchanged messages' sequence, etc., without focusing on additional network-related information or user-related context-information. Several claimed patents are based on network related-information such as bandwidth, delay etc. A smaller number of these take into account user-related information such as application information, prediction of location, etc. None of them, however, correlates application usage, past location, mobility patterns, device characteristics and capabilities in order to finally build a generic user profile and use it in order to realize resource mapping from a holistic network point of view.

2.2. KDD tools

Methodologies and tools for KDD are divided into three categories: unsupervised, supervised and semi-supervised. *Unsupervised* KDD methods assume the existence of pattern(s) in data which they try to unveil (e.g. identify *clusters* of similar observations). *Supervised* KDD methods on the other hand focus on learning existing pattern(s) from available data (i.e. training set) and then use them in order to classify previously unknown observations (e.g. assign a new object to a set of predefined *classes*). Finally, *Semi-Supervised* KDD bridges the two aforementioned genres by attempting to identify pattern(s) in datasets (like Unsupervised KDD) using information provided by a limited training set. Before delving into the details of our framework we present a number of KDD algorithms which we employ in our work.

2.2.1. k-Means clustering

k-Means [67] is a partitioning clustering algorithm used for many unsupervised learning tasks. k-means tries to separate samples in k groups of equal variance by minimizing the sum of intra-cluster distances. The algorithm essentially minimizes the objective function: $\sum_{i=1}^{k}\sum_{x\in C_i}||x-m_i||^2$ where x is an object from the dataset and m_i the centroid of cluster C_i. k-Means is often referred to as Lloyd's algorithm [68].

2.2.2. Spectral clustering

Spectral clustering operates on the first k eigenvectors derived from the decomposition of the data graph Laplacian and applies k-Means on the projected dataset in order to derive the clusters. The algorithm operates on a large graph defined by the data similarities which may be a full graph, a k-nn graph (only the k-nearest neighbors of each object are retained) or an e-neighbourhood graph (only the points within range e are retained). The graph's adjacency matrix $\boldsymbol{W}$ is used for the definition of $\boldsymbol{D}$, a diagonal matrix with $D_{i,i}=\sum_{j=1}^{n}W_{i,j}$. Using $\boldsymbol{D}$ and $\boldsymbol{W}$ we define the graph Laplacian $\boldsymbol{L}$:

- Unnormalized Graph Laplacian: $\boldsymbol{L}= D\text{-}\boldsymbol{W}$
- Normalized Graph Laplacian: $\boldsymbol{L_{sym}} = \boldsymbol{I}\text{-}\boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}$
- Normalized Graph Laplacian: $\boldsymbol{L_{rw}} = \boldsymbol{I} - \boldsymbol{D}^{-1}\,\boldsymbol{W}$

The first k eigenvectors of $\boldsymbol{L}$ comprise the embedding of the original dataset in R^k. The eigenvalues $\boldsymbol{\Sigma}$ of $\boldsymbol{L}$ are used for identifying the value of k; the multiplicity of the zero eigenvalue provides an approximation of k since each zero eigenvalue denotes a disconnected component of the graph (i.e. a cluster). If the graph is connected (i.e. $\sigma_i \neq 0 \forall i$) but data enjoy a clustered structure, we expect a large gap in the eigenvalues; assuming that $\sigma_i \le \sigma_{i+1}\ \forall\ i$, then $k{=}i$ when $|\sigma_{i+1}-\sigma_i| \gg |\sigma_{j+1}-\sigma_j|\ \forall\ i \neq j$. The latter is known as the "eigengap heuristic" . k-Means is applied in the end in order to discretize the solution and assign labels to clusters [69,70].

2.2.3. Naïve Bayes

Bayesian methods are a set of supervised learning algorithms based on Bayes' theorem [67]. Bayesian methods are called "Naive" due to the fact that they operate under the salient assumption that class membership depends on only one variable rather than their combination. From a practical perspective, calculations are fast and simple; given a training set with C classes, we compute $P(C_j),\ j=1\ldots C$ for all classes of the set. Then, by considering the "naïve" approach we quantify $P(x_i|C_j)$ for all instances of class C_j and all values of variable x_j from the set of observations.

2.2.4. ID3 decision tree

Decision Trees are powerful learning mechanisms used for classification and prediction. The ID3 algorithm [67] is the most known and widely used due to its simplicity and effectiveness. ID3 iterates over the dataset and divides it along variables taking into account their entropy. Specifically, on each iteration, it goes through remaining variables, calculates their entropy and uses the variable

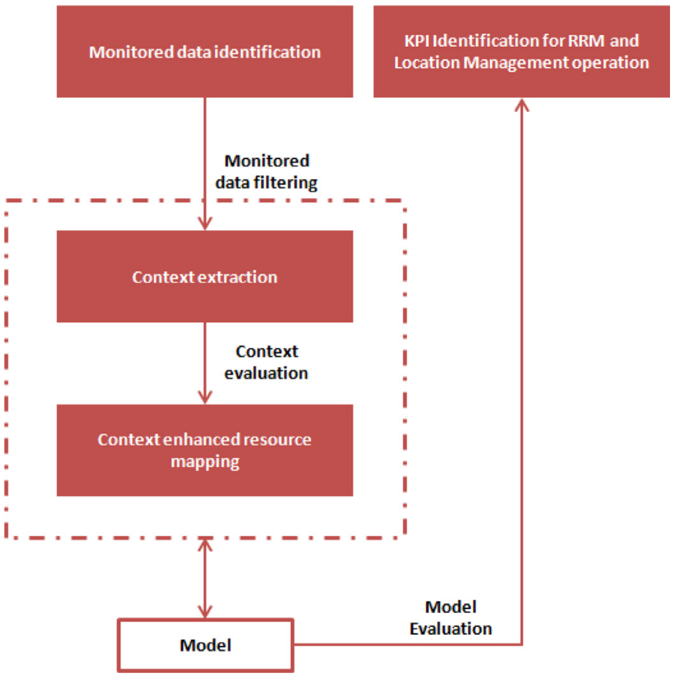

Fig. 1. Methodology for knowledge extraction and RRM operation via RAT selection.

with the smallest entropy in order to split the dataset. The algorithm continues on the remaining attributes until either all variables are used or there are no more observations to divide (e.g. empty dataset or all remaining data belong to the same class)

2.3. Motivating discussion

The comprehensive overview of the various solutions proposed by the researcher community, leads us to the conclusion that most approaches are either too simple to implement but achieve sub-optimal solutions, or provide significant improvements but their complexity or the burden placed on the network components renders them unattractive for a real deployment by the operators. It is therefore evident that existing solutions need to cover a larger gap in order for RRM mechanisms to be able to efficiently and realistically support the real needs and requirements of 5G networks, with one of the primary challenges being to deal with the constantly increasing number of mobile users and bandwidth-intensive services [12] via effective and efficient network planning.

An important issue is related to the additional information that needs to be exploited by the network. A novel, overarching framework, on top of all three control schemes (i.e., cell (re)selection, call admission control (CAC) and handover) should be able to take advantage of multiple sources of information (UE or network oriented) and extract from it additional knowledge. The new scheme that we envisage collects information about users, services, terminals and network conditions and –based on offline processing and knowledge extraction– categorizes the UEs according to their behavior. From an architectural perspective, the derived models are passed to the HSS and to the serving MME of a terminal so as to be exploited during the cell selection, CAC and handover processes.

3. The context extraction and profiling engine

Fig. 1 describes the methodology for the knowledge extraction and the subsequent enhancement of the afore-described mobility control schemes.

The first step is to identify the data that should be monitored, collected and processed as well as the Key Performance Indicators (KPIs) which will be used in the end in order to assess the effectiveness and efficiency of the derived model. At least four data types and associated KPIs should be selected and used, namely: network operation data, user behavior information, terminal capabilities and service data. Afterwards we proceed with the evaluation of the contextual information which entails the extraction of data (e.g. selection of measurement according to location, time etc) and the derivation a model which will be used in order to create enhanced cell (re)selection, CAC and handover mechanisms. The model is finally evaluated against the initially defined set of KPIs.

It should be noted that CEPE is not an algorithmic solution but a KDD framework focusing on the exploitation of available contextual information in order to dynamically identify profiles and associate them with sets of rules which upon application can ameliorate the overall network operation (i.e. provide a more efficient RAT/layer mapping of UEs).

In section III.A we provide a high level description of CEPE; the unsupervised version of our framework appears in section III.B while its supervised counterpart in section III.0. The description includes algorithmic details and computational complexity considerations. Model evaluation and querying is presented in sections III.0 and III.E respectively.

3.1. High level description

For simplicity reasons, we assume that each observation derived from our ecosystem resembles a row in a log file consisting of all the monitored parameters. For example, assume that a *User* with his *Device* starts at *Time* t_0 consuming a *Service* through a *Network* technology. *User* refers to the unique identifier of the specific UE, *Device* to the specific type of equipment that is used along with the device capabilities (CPU, monitor, etc.) and the *Service* relates to the type of the session that is active when the particular information is logged (i.e., VoIP call, browsing, Video streaming, etc.). Finally, the *Network* refers to the type of the RAT, as well as the cell layer via which, the specific service is being consumed (e.g., Wi-Fi, LTE femto-cell, macro cell, etc.). Consider now that we take a "snapshot" of the system every *t* seconds. Thus our log input will be made up of numerous rows that look like:

$User \bowtie Device \bowtie Service \bowtie Network@Time$

Obviously, using the available information, we can derive additional parameters and augment our model. For example:

- *Uplink Peak User Throughput:* find the max value within a specific time frame
- *Downlink Peak User Throughput:* find the max value within a specific time frame
- *Uplink Delay:* calculate uplink delay within a specific time frame
- *Downlink Delay:* calculate downlink delay within a specific time frame

Similarly we can quantify any required indicator and augment our input data:

$\{User \bowtie Device \bowtie Service \bowtie Network@Time,\ \mathrm{KPI}_1,\ \ldots,\ KPI_n\}$

Note here that the level of granularity can change by selecting different time periods or a different entity. For example we may ignore the *User* and *Device* axes of our data model while aggregate records per Week and evaluate the effect of *Service* consumption on our *Network*:

$\{Service \bowtie Network\ @\ Time_{Week},\ \mathrm{KPI}_1,\ \mathrm{KPI}_2,\ \ldots,\ KPI_n\}$

During the design of CEPE we took into consideration the huge amount of available information as well as the need to formulate and derive various groups and associations, either apparent or latent, offline or at near real time. CEPE therefore operates on two levels (i) Macroscopic/ Horizontal where each observation is

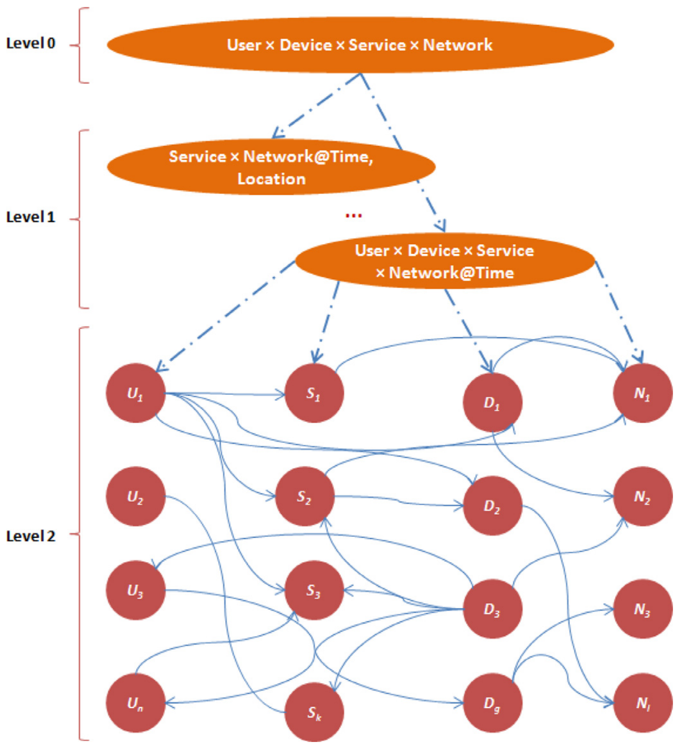

Fig. 2. Schematic depiction of the envisaged KDD. U_i, S_i, D_i and N_i depict distinct user, service, device and network groups.

treated as a whole and (ii) Microscopic/ Vertical where all operations are applied on the elements defining each observation.[1] This two-level approach facilitates the application of CEPE on large data collections while in parallel –since a significant portion of the information remains static (e.g. device type) – speeds up classification. From a high level, methodological perspective, the proposed framework builds a is implemented in three steps (refer to Fig. 2):

- Data are aggregated from the identified data sources and formulate the dataset to be processed (Level 0)
- The dataset is broken down into subsets in order to derive sets of similar observations (Level 1, e.g. similar time period, similar geographic area etc).
- Each subset is broken down into entity specific chunks following a simple disaggregation approach; each observation is broken down into its constituting entities (Level 2, e.g. *User, Device, Network* and *Service* sets).

Similar observations per entity are grouped together (e.g. ***User*** = {$User_1$, $User_2$, ..., $User_N$}). We assume that each group defines a node in a graph which is connected with another entity node via an edge with weight $w_{i,j}$ where i and j denote groups of different entities (e.g. $User_i$ and $Service_j$ are connected with an edge of weigh $tw_{i,j}$).

The next step is to answer questions; for example, how to optimally assign a *User* to a RAT (e.g. cell id and location) taking into account contextual information. A naïve usage of the model for this purpose would be consider the *User* entity, identify the *User*'s group, traverse the graph and find the most proper *User-Service-Device-Network* path (e.g. a path that maximizes the sum of weights). As we will see in the following paragraphs, we exploit this graph-traversal approach but apply a more elaborate scheme.

[1] Recall that an observation comprises the concatenation of instantiations of different participating entities (e.g. User x Device x Service Status @ Time).

Table 1
Employed mathematical notation.

Symbol	Explanation
$\boldsymbol{x}$	*Row vector x*
x_i	*The ith coordinate of* $\boldsymbol{x}$
x	*Variable x*
x^t	*The value of variable x at time t*
$\boldsymbol{X}$	*Matrix X with observations as rows*
$\boldsymbol{x}^T$ or $\boldsymbol{X}^T$	*Transpose of* $\boldsymbol{x}$, $\boldsymbol{X}$
X_{ij}	*The value of cell i,j of* $\boldsymbol{X}$
$\mathbf{X}_{:,j}$ or $\mathbf{X}_{j,:}$	The jth column/row of matrix $\mathbf{X}$
$\mathbf{X}_{1:n,:}$ or $\mathbf{X}_{:,1:n}$	All rows/columns of $\mathbf{X}$ from index 1 to n
$\mathbf{\|X\|}$	The number of observations (rows) in $\mathbf{X}$
U, D, S, N	The matrices of user, device, service and network observations respectively

Building and updating this model takes place offline, since it is time consuming. The updates are based on a sliding window approach where a set of observations is replaced by a new one. This action is executed periodically, in the data-warehouse of the network operator, since it implies rebuilding the models.

Searching on the other hand can take place in real-time. *Network* and *Device* nodes are few (devices can be roughly categorized into maximum 10 groups; types of networks even less) while user nodes are also limited by the input *User*. The added value of CEPE is that it is generic in the sense that it can be applied to the whole dataset, a subset of the dataset or a time-projection of it. This means that you can obtain different models for different geographic areas, time slots etc.

CEPE can be applied in either a supervised or an unsupervised context thus leading to different results. In the unsupervised CEPE, we assume that there are some groups in our data which we attempt to identify. On the contrary, in the supervised CEPE, we know in advance the groups (i.e. classes) and attempt to define a model that best describes them so as to be able in the future to categorize previously unobserved instances.

Evidently, everything comes at a cost; in the unsupervised case we gain flexibility and adaptability but the approach is prone to ill-defined and noisy data. In the supervised case we gain robustness but lose adaptability. For example, consider the case where a network operator identifies a new user group based on accumulated charging data records (CDRs) and customer information (CRM). Using this information, he sets up a campaign (e.g. offers free MBs to low spenders), which in turn results in the definition of a new group (e.g. low spenders who exhibit high-spender characteristics for a certain amount of time). Supervised CEPE will successfully categorize users in the initial group set but will fail to identify the new one. The unsupervised CEPE will exhibit worse performance in the first case –identification of profiles– but will find out that a new group has emerged when enough data are accumulated.

In the rest of the paper, we assume that all entities are represented as a high dimensional vector residing in R^n. Before delving into details, we provide a table (Table 1) summarizing the basic mathematical notation we employ. Note that our vectorization assumption renders the algorithm data agnostic and capable of accommodating changes in the underlying data model. Indeed, any change in the raw data necessitates an update in the mapping function (i.e. the function which will map an entity –for example a *User*– as a high dimensional point) leaving CEPE unchanged.

3.2. CEPE unsupervised version

The unsupervised version of CEPE operates under the salient assumption that there is some latent structure in the data collection (i.e. groups/clusters of similar observations) which attempts to unveil and formalize it through rules. The algorithm appears in Table 2.

Table 2
The unsupervised version of CEPE approach.

Step	*function: unsupervised_CEPE*	
	Input parameters	***X:*** The dataset where rows are observations and columns are monitored variables
	u, d, s, n←number of user/ device/ service/ network-related variables in observations	
1.	$\boldsymbol{U} \leftarrow \boldsymbol{X}_{:,\,1:\,u}$ $\boldsymbol{D} \leftarrow \boldsymbol{X}_{:,u+1:u+d}$ $\boldsymbol{S} \leftarrow \boldsymbol{X}_{:,u+d+1:u+d+s}$ $\boldsymbol{N} \leftarrow \boldsymbol{X}_{:,u+d+s+1:\lvert \boldsymbol{X}^T \rvert}$	Divide the dataset into entity specific chunks.
2.	**ul,Uc** ← spectral_clustering(**U**) ur ← decision_tree(**U,ul**) **dl,Dc** ← spectral_clustering(**D**) dr ← decision_tree(**D,dl**)	Run spectral clustering and decision tree classification for all derived matrix chunks. Spectral clustering facilitates the identification of clusters (i.e. groups of similar observations) while the decision tree classifier extracts a rule-set which will be used for the classification/mapping of new observations to the identified clusters.
	sl,*Sc* ← spectral_clustering(**S**) sr ← decision_tree(**S,sl**)	**ul** denotes the label vector of user clustering (e.g. 'user_group_x', ur the classification rules (e.g. age> 18 ^ income< 50 K→user_group_x) and **Uc** the corresponding centroids (e.g. average age 22.5, average income 35 K).
	nl,Nc ← spectral_clustering(**N**) nr ← decision_tree(**N,nl**)	
3.	**v** ←distinct_classes(**ul** ∪ **dl** ∪ **sl** ∪ **nl**)	Merge all label vector and identify the distinct elements (i.e. cluster labels) that will form the nodes of the graph. Store them into vector **v.**
4.	$E_{i,\,j} \leftarrow P(v_i \mid v_j),\ \forall\, i \neq j$	Calculate the adjacency matrix
5.	**G** ← full_graph(**v,E**)	Generate the graph
	Output parameters	**G**: The graph ur, dr, sr, nr: The set of rules **ul, dl, sl, nl**: The set of labels **Uc, Dc, *Sc*, Nc**: The set of centroids

We assume that the dataset is split into subsets in accordance with the variables in consideration. An intuitive selection could be the time and location axes (e.g. weekdays, 9:00–12:00 around coordinates {(x,y),(z,a)}) however this is indicative since the procedure can be applied along any data dimension. This step corresponds to descending from Level 0 to Level 1 as depicted in Fig. 2.

Thereinafter, we break down every derived subset into entity specific data-chunks following a simple disaggregation approach; each observation is broken down into its constituting entities (e.g. User, Device, Network and Service chunks as depicted in Step 1 of Table 2).

The next step entails the application of spectral clustering on the identified chunks and the derivation of entity-specific clusters (Step 2). We assume that the derived membership is correct and generate a knowledge model using a decision tree classifier. The tree model facilitates abstraction and generalization; the rules will help us categorize new instances faster without maintaining the whole dataset in memory or constantly updating the spectral decomposition.

Each distinct cluster label defines a node in a graph (Step 3), which is connected with another entity node via a weighted edge w_{ij}. The simplest way to weight edges is the use of conditional probabilities; for example, when connecting a user cluster with a service cluster, we can define a weight:

$$w_{i,j} = \frac{\textit{times that a user of this cluster consumed a service of that cluster}}{\textit{number of services}}.$$

The possible paths that traverse all entities (i.e. from a given *User* node to any *Network* node traversing all other entities) correspond to the different profiles (i.e. the different combinations of User – Device – Service – Network labels that CEPE identified in the data collection).

The spectral clustering algorithm appears in Table 3; at first we calculate the graph Laplacian L (Step 1) by taking into account the data pairwise similarities. Afterwards we derive its eigen decomposition (Step 2) and apply the eigengap heuristic on the matrix of eigenvalues $\mathbf{\Sigma}$ ($\mathbf{\Sigma}$ is a diagonal matrix with eigenvalues in descending order along its main diagonal – i.e. $\sigma_{ii}\neq 0$ and $\sigma_{ij}=0$ when $i\neq j$).

Spectral clustering is both a clustering and a dimensionality reduction algorithm. The new dataset is obtained by retaining the first k eigenvectors of $\mathbf{E}$ (Step 4) and embeds the original observations from R^n to R^k where n is the number of variables. We employ the k-means algorithm in order to discretize the result (i.e. derive the clusters from $\mathbf{X_{new}}$). Finally, taking into account the cluster membership information we calculate the centroids on the original matrix $\mathbf{X}$ (Step 6).

The computational complexity of the procedure is primarily dominated by the application of spectral clustering on matrixes **U, D, S** and **N** as well as the formation of graph **G** thus is upper bounded by $O(|\mathbf{v}|^2+|\mathbf{U}|^3+|S|^3+|\mathbf{D}|^3+|\mathbf{N}|^3)$. Memory requirements are upper bounded by $O(|\mathbf{v}|^2+|\mathbf{U}|^2+|S|^2+|\mathbf{D}|^2+|\mathbf{N}|^2)$ due to the storage of the graph **G** and the eigen decomposition of **U, D, S** and **N**.

3.3. CEPE supervised version

The key differentiating factor of the supervised case with respect to the unsupervised one is that we know in advance how we should break the data chunks into distinct clusters. The latter is due to the fact that we are aware of the underlying structure of our data collection.

The formal description of the algorithm appears in Table 4. Again we assume that the dataset is split into subsets in accordance with the variables in consideration and we break down every derived subset into entity specific data-chunks as depicted in Step 1.

As soon as the preprocessing is performed, data is fed to a classifier (Step 2). Note that in general we can apply any type of clas-

Table 3
The spectral clustering algorithm.

Step	*function: spectral_clustering*	
	Input parameters	***X:*** The dataset where rows are observations and columns are variables
1.	$\boldsymbol{L} \leftarrow \boldsymbol{D} - \boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}$	Calculate the un-normalized graph Laplacian $\mathbf{L}$[a]. $\mathbf{D}$ is the degree matrix and $\mathbf{XX^T}$ the observations' pairwise similarities matrix according to their internal product (i.e. $\sum_{i=1}^{\lvert \boldsymbol{x} \rvert} x_i y_i$ where **x,y** are observations –i.e. rows– of **X**).
2.	$\boldsymbol{E}, \boldsymbol{\Sigma} \leftarrow eig(\boldsymbol{L})$	Derive the eigen-decomposition of **L**
3.	k ← eigengap (**Σ**)	Use the eigengap heuristic and find the number of clusters k.
4.	$\boldsymbol{X_{new}} \leftarrow \boldsymbol{E}_{:,\,1:\,k}$	Derive the new dataset.
5.	**labels** ← k-Means($\mathbf{X_{new}}$,k)	Run k-means clustering in order to discretize the result.
6.	**C** ← centroids(**X,labels**)	Calculate the centroids on the original matrix **X**.
	Output parameters	**labels**: The labeling scheme as derived from k-means **C**: The centroids calculated on the original matrix X.

[a] The normalized graph Laplacian can also be used as well as any scheme (e.g. kernel similarity, k-NN search etc) for the definition of the pairwise similarities matrix W. For simplicity reasons we present only the un-normalized graph Laplacian case together with the internal product similarity computation.

Table 4
The supervised version of CEPE approach.

Step	*function: supervised_CEPE*	
	Input parameters	***X***: The dataset where rows are observations and columns are monitored variables *ul, sl, dl, nl*: The labeling scheme
	u, d, s, n←number of user/ device/ service/ network-related variables in observations	
1.	$\boldsymbol{U} \leftarrow \boldsymbol{X}_{:,\,1:\,u}$ $\boldsymbol{D} \leftarrow \boldsymbol{X}_{:,u+1:u+d}$ $\boldsymbol{S} \leftarrow \boldsymbol{X}_{:,u+d+1:u+d+s}$ $\boldsymbol{N} \leftarrow \boldsymbol{X}_{:,u+d+s+1:\lvert \boldsymbol{X}^{\boldsymbol{T}} \rvert}$	Divide the dataset into entity specific chunks.
2.	*ur ← decision_tree(**U**,**ul**)*	Run decision tree classification for all derived matrix chunks. ur denotes the classification rules (e.g. age> 18 ^ income< 50 K → user_group_x).
	*dr ← decision_tree(**D**,**dl**)* *sr ← decision_tree(**S**,**sl**)* *nr ← decision_tree(**N**,**nl**)*	
3.	***v** ← distinct_classes(**ul** ∪ **dl** ∪ **sl** ∪ **nl**)*	Merge all label vector and identify the distinct elements (i.e. cluster labels) that will form the nodes of the graph. Store them into vector **v.**
4.	$E_{i,\,j} \leftarrow P(v_i \vert v_j),\ \forall\, i \neq j$	Calculate the adjacency matrix
5.	***G** ← full_graph(**v,E**)*	Generate the graph
	Output parameters	***G***: The graph *ur, dr, sr, nr*: The sets of classification rules

sification algorithm. A Tree Classifier –e.g. the ID3 we used in the experiments & the unsupervised case– will produce a set of classification rules (i.e. *X & Y→Z*). A Lazy Learner –e.g. kNN, [67]– will construct a tree and then identify for each incoming observation its *k* closest instances. A multiclass SVM classifier – [67]– will produce, for each pair of classes, a set of support vectors which will define a linear equation that optimally separates the two classes. Evidently, any classifier is directly applicable without affecting the methodology.

The rest of the procedure is identical to the unsupervised CEPE. Conceptually, the approach is the same; the key differentiating factor, –which eventually provides better discrimination results–, is the existence of the correct labeling scheme, which significantly enhances and speeds up the procedure.

The computational complexity of the procedure is primarily dominated by the application of the decision tree classifier on matrixes **U, D, S** and **N** as well as the formation of graph **G** thus is upper bounded by $O(\lvert\mathbf{v}\rvert^2+(u\lvert\mathbf{U}\rvert)^2+(s\lvert\mathbf{S}\rvert)^2+(d\lvert\mathbf{D}\rvert)^2+(n\lvert\mathbf{N}\rvert)^2)$ where u, d, s, n are the number of user/ device/ service/ network-related variables in observations and |**U**|, |**S**|, |**D**| and |**N**| the number of respective observations. Memory requirements are upper bounded by $O(\lvert\mathbf{v}\rvert^2+\lvert\mathbf{X}\rvert)$ due to the storage of the graph **G** and original data matrix.

3.4. Rules extraction and feedback loop

Until this point, CEPE description has focused solely on building and maintaining a knowledge base containing groups and associations between users, devices (e.g. type, capabilities, battery status, mobility, charging etc), RATs (e.g. type, cell-id, location) and Services. We have stated however, that the goal is to map UEs to RATs and proposed to do so via a set of rules. In the context of this paragraph, we will discuss the derivation of these rules from our knowledge model and their evaluation.

A first straightforward approach could be the manual extraction and evaluation of rules. A human expert (e.g. network administrator) identifies the various groups, studies their properties (e.g. group *A* is made up of prepaid subscribers that perform *x* top-ups a month and communicate primarily with SMS) and then derives the combinations that according to his expertise will optimize network operation. For example, consider an unsupervised case were we have identified the following profiles:

- User Group X: User Group 19–29, Prepaid subscription, 2GB available for data, when at 20% of credit, service consumption and calls drop sharply by 50%, moving at high velocity
- Service Group Y: Video Streaming (i.e. YouTube), VoIP Services (i.e. Skype) take up more than 80% of his time
- Device Group Z: Samsung Galaxy S4

A human administrator would probably come up with a rule that optimizes the QoE of the user class and avoids excessive network signaling, e.g., reduce the large number of HOs occurring due to high user velocity. A probable rule could be the following:

- User Group X ^ Service Group Y ^ Device Group Z → Macro Cell

This approach is plausible for an average number of groups/classes per entity (e.g. less than 10 per entity) considering that numerous cases can be grouped together. But obviously, if we need finer granularity (i.e. higher level of detail per class) we need to come up with a semi-supervised or a totally unsupervised approach for rules derivation.

Assuming that the network is configured to take optimal decisions most of the time, we can autonomously generate a set of rules, which upon application can ameliorate network conditions. The derivation is based on the graph constructed in the final step of CEPE. Recall that every node of the graph represents a distinct class of a particular entity (i.e. U_1 in Fig. 2 corresponds to a class label of entity *User*) and is connected with other nodes via weighted edges with edge weights depicting the probability of having instances of both classes on the same path.

We can employ two distinct strategies in order to derive the rule-set; directly apply Bayesian logic (i.e. Naïve Bayes classification), find all possible rules and rank them according to their score or alternatively identify the paths that traverse all class types (i.e. *User, Device, Service* and *Network*) and rank them according to the sum of weights.

This way, the case with the highest probability (i.e. appears most of time) is the one applied as a rule in similar situations. However, such an approach is prone to mis-configurations; a correct but rare decision will be ignored and never be applied even when it should be. Evidently, in order to employ this procedure, we need a kind of feedback loop that will promote correct rules and degrade those invalid.

Recall that in the beginning we proposed the extraction of a set of KPIs against which CEPE performance will be evaluated; depending on the induced amelioration or deterioration on these KPIs we can create a score for each rule in the form $\frac{1}{n}\sum_{i=1}^{n} w_i p_i$where w_i is an optional weight – importance – for the *i*th KPI defined by an administrator, p_i the percentage change of the *i*th KPI due to the application of the rule in question and *n* the number of evaluated KPIs.

The approach is depicted in Tables 5 and 6. We assume that the network initiates with the rules derived according to the graph traversal plan and operates with these rules for a given time period *t*; afterwards, a set of predefined KPIs *x* is evaluated per user against their counterparts during operation period *t-1*.

For each set of *n* KPIs we calculate $\frac{1}{n}\sum_{i=1}^{n}\frac{(\boldsymbol{x}_i^{t-1}-\boldsymbol{x}_i^{t})}{\boldsymbol{x}_i^{t-1}}$, where $p_i = \frac{(\boldsymbol{x}_i^{t-1}-\boldsymbol{x}_i^{t})}{\boldsymbol{x}_i^{t-1}}$, $\boldsymbol{x}_i^{t}$ the value of the i-th KPI obtained during trial period *t* and $\boldsymbol{x}_i^{t-1}$ the same value during the trial period *t-1* (or in the case that $t = 1$ when the system run without CEPE). Depending on the importance of each KPI we can adjust the sum by including weights from a set $\boldsymbol{w}$ thus obtaining the final score for a given rule. Finally we sum up the individual scores per user and obtain a holistic value for each rule or set of rules.

3.5. Querying a CEPE defined model

The algorithmic solutions presented in the previous sections create the knowledge base and associate rules set, upon which decisions will take place. In the context of this paragraph we will focus on the decision step; given a CEPE model and an observation, how to best assign it to a specific class and which rule should CEPE advice for invocation?

Recall that we have essentially structured a set of profiles and a set of rules so given an observation we want to identify the optimal set of classes (i.e. profile), on which it should be mapped and the proper rule to invoke.

Querying can take place either real-time or offline. Real-time search means that the required parameters will be periodically transmitted from the UE to the network, which in turn will feed them to the model and derive the classification of the observation.

It now becomes apparent that the vertical division step is extremely helpful since static information (e.g. device capabilities, user preferences etc.) will not change thus classification will take place only once for the updated entities. As soon as the observation has been properly classified, we match it with the rules and apply the one with the highest rank. In case of multiple matching rules we can randomly select and apply one since the subsequent evaluation step will assess its correctness.

Offline querying on the other hand assumes that user behavior exhibits strong periodicity in terms of time and location. Therefore the user profile and service profiles will not change over time enabling the exploitation of previous decisions. The approach appears in Table 7.

3.6. Experimental evaluation

In order to assess the validity and viability of our approach, we performed extensive experiments using the NS3 network simulator [73] and custom Python modules. We considered a usage scenario from the METIS project [71,72] which we implemented in NS3 and evaluated the application of CEPE (implemented in Python) on the case of handovers.

Through the experiments we attempted to replicate –to the best possible extent– a real life situation. Towards this end we conducted an extensive literature review that covered a large number of aspects like mobility speed, energy consumption patterns, service usage patterns, etc. We present our findings and configuration in the following.

3.6.1. Experimentation scenario and setup

The scenario considers a large shopping mall with high density of customers and service staff (essentially, an established 5G scenario, i.e., Ultra Dense Environment). A typical setting for a future extended rich communication environment, involves “traditional” radio networks and wireless sensor networks, where customers access mobile broadband communication services while they are directly addressed by personalized location-based services of the

Table 5
Evaluate a rule-set using feedback from all subscribers.

Step	*function:* evaluate_rule_set	
	Input parameters	r^t: A vector of matrixes containing as elements one matrix per KPI for evaluation time *t*. r^{t-1}: A vector of matrixes containing as elements one matrix per KPI for evaluation time *t-1*. *w*: A vector containing the weights for each KPI.
1.	*r* ← *number of elements in* $\boldsymbol{r^t}$	Find the number of rules included for evaluation
2.	*score* ← *0,* $\boldsymbol{s}$ ← Ø	
3.	*for i =1:1:r* $\boldsymbol{s}_i$ = *evaluate_rule*($\boldsymbol{r}_i^t, \boldsymbol{r}_i^{t-1}, \boldsymbol{w}$)*score* += $\boldsymbol{s_i}$	For all rules in this evaluation run, calculate the induced amelioration/ degradation percentage and add it to the overall score.
4.	*score* = $\frac{1}{r}$*score*	Normalize the score taking into account the number of rules
	Output parameters	*score*: The score obtained for the particular rule set $\boldsymbol{s}$: A vector containing the scores for all rules. The i-th element of $\boldsymbol{s}$ contains the score of the i-th rule.

Table 6
Evaluating a single rule using feedback from all subscribers.

Step	*function: evaluate_rule*	
	Input parameters	$\boldsymbol{X^t}$: A matrix containing as columns the values of the KPIs at time t for all users. $\boldsymbol{X^{t-1}}$: A vector containing as columns the values of the KPIs at time t-1 for all users. ***w:*** A vector containing the weights for each KPI.
1.	*u* ← *number of rows of* $\boldsymbol{X}^t$	Find the number of users and KPIs included in this evaluation
	n ← *number of columns of* $\boldsymbol{X}^t$	
2.	*score* ← *0*	
3.	*for i = 1:1:u*	
	score+ = $\frac{1}{n}\sum_{j=1}^{n} w_j \frac{(\boldsymbol{X}_{i,j}^{t-1}-\boldsymbol{X}_{i,j}^{t})}{\boldsymbol{X}_{i,j}^{t-1}}$	For all users in this evaluation run, calculate the induced amelioration/ degradation percentage and add it to the overall score.
4.	*score* = $\frac{1}{u}$*score*	Normalize the score taking into account the number of users
	Output parameters	*score*: The score obtained for the particular rule

Table 7
Querying a CEPE model.

Step	*function: query_CEPE*	
	Input parameters	$\boldsymbol{o}$: An observation ***ur, dr, sr:*** The sets of classification rules as derived from the decision tree classifiers of either supervised of unsupervised CEPE. *online/offline*: A flag signifying whether the procedure will run real-time of offline ***N:*** The set of network mapping rules as derived from the graph traversal process
	u, d, s ← *number of user/ device/ service variables in the given observation*	
1.	$\boldsymbol{u} \leftarrow \boldsymbol{o}_{1:u}$ $\boldsymbol{d} \leftarrow \boldsymbol{o}_{u+1:u+d}$ $\boldsymbol{s} \leftarrow \boldsymbol{o}_{u+d+1:\lvert\boldsymbol{o}^T\rvert}$ *if online:*	Divide the observation into entities
2.	*ul* ← *classify(**u**,**ur**)*	Run the classifier for all derived observation chunks and attribute them the most fitting label (ul, dl, sl).
	dl ← *classify(**d**,**dr**)* *sl* ← *classify (**s**,**sr**)* *else:*	
3.	*ul, dl, sl* ← *get_historic(**u, d, s**)*	Search past observations
4.	***R*** ← *get_rules(ul, dl, sl, **N**)*	The set of rules accompanied by their score as derived by the graph traversal process.
	Output parameters	R_1: The rule with the highest score

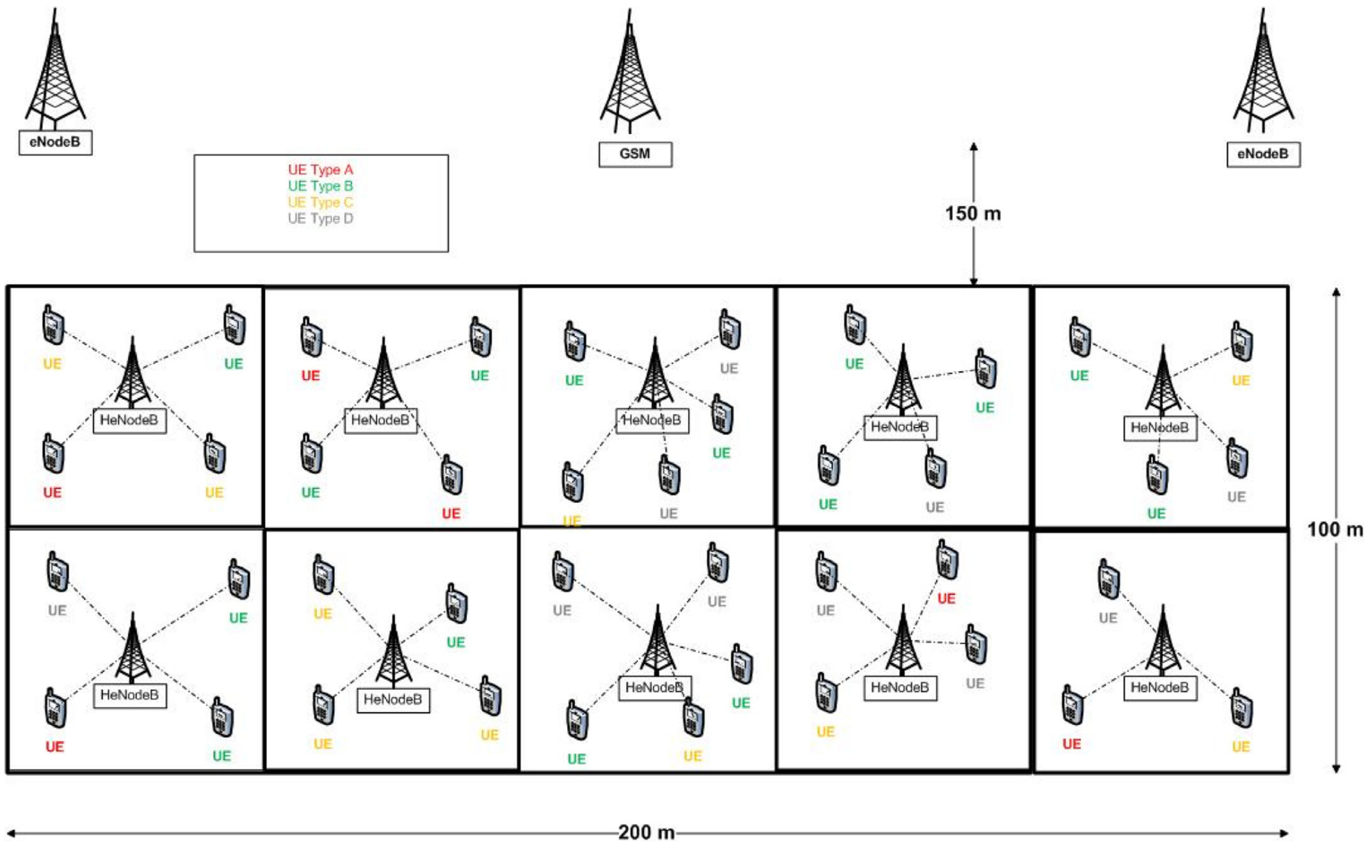

Fig. 3. Simulation topology.

Table 8
NS-3 configuration.

NS3 network node	Tx power (dBm) [74]	Downlink (DL) Earfcn (MHz) [74]	Bandwidth (RBs) [74,75]	Antenna type [74]
Macro cell	35	2120	50 (10 MHz)	Parabolic, 15 dBi
Femto cell	20	2120	15 (3 MHz)	Isotropic
GSM	35	2120	15 (3 MHz)	Parabolic, 15 dBi
UE	20	–	–	Isotropic
Macro cell	35	2120	50 (10 MHz)	Parabolic, 15 dBi

shopping environment. Overall, the network deployment allows seamless handling of services across different domains, e.g. mobile/fixed network operators, real estate/shop owners, application providers. Based on this description, we use the NS3 and model a single floor, 200×100 m building, containing 10 rooms, with an LTE Femto cell placed in each of them. Outside, two LTE eNBs are placed, 150 m north of the mall with Inter-Site Distance (ISD) equal to 200 m, and a GSM cell between them (at equal distance from the eNBs). Fig. 3 depicts the considered simulation topology.

Our simulation scenario is based on 3GPP Specifications [74] and [75]. In details, the transmission mode is SISO (Single Input Single Output); the handover algorithm is the A2A4 RSRQ-based[2] and the scheduler is the NS-3 implementation of the Proportional Fair MAC scheduler [75]. We use the Hybrid Buildings Propagation Loss Model for path loss implemented in NS3 with Internal Wall Loss at 10.0 dB Shadow, Sigma Indoor at 10.0 dB [74]. The network node configuration appears in Table 8. Services are implemented using NS3's UDP client-server application model and the desired data rates are achieved through configuration of the packet size and the inter-packet interval parameters. The service schedule for every user is pseudo-randomly generated at the beginning; as the simulation progresses they affect and are affected by the battery state and the charging status. Service parameters appear in Table 9 Each time one of the services is triggered according to the service schedule mentioned above, a constant bit rate traffic model is generated with the respective duration; the traffic is between the clients (UEs) and a remote host, while our measurements concern only the part of the access network.

Every user follows a mobility model comprising a) the velocity and b) the path pattern (linear, random, etc.). The mobility model may change during simulation. Every 4 min, each user randomly selects a model; additionally, when a UE has moved 30 m towards any direction, it randomly selects another direction to move next. The considered mobility models are: Stationary Mobility (0 m/s – 0.8 m/s) where customers move very slow or remain at their position; Low Mobility (0.8 m/s – 1.4 m/s) where customers move with a slow or average pace inside the mall; and Medium Mobility (2 m/s +/−0.6) where customers walk fast inside the mall. Moreover, every user has a charging level denoted as Bronze, Silver or Gold (randomly assigned based on a uniform distribution), emulating the data capacity of his subscription. Bronze users have a maximum of 500 MB to spend on data services (the initial value is randomly generated between 40–500 MB), Silver users a max-

[2] https://www.nsnam.org/docs/models/html/lte-design.html#fig-lte-legacy-handover-algorithm.

Table 9
Service parameters used in simulation.

Service Type	characteristic	Value	Comments
Short duration voice	Duration	100 s (+/−10)	Average call duration is 1.8′ [76]
	Rate	13 kbps UL & DL	Average rate is 12.65 kbps [77,78]
Long duration voice	Duration	240 s (+/−20)	
	Rate	13 kbps UL & DL	
Web data	Duration	4 s (+/−2)	Average web page access session duration is 4.2 seconds [79]
	Rate	1.6 MB DL (+/−500 KB)	Average web page size is 1.6MBs [80]
FTP data	Duration	9 s (+/−2)	Average session for file download is 9.8 seconds for 3MBs file [79]
	Rate	3 MB DL (+/−60 KB)	
Video stream data	Duration	240 s (+/−30)	Average YouTube video duration is 4.12′ [84] Average DL speed 443 kbps [82]
	Rate	13.5 MB DL (+/−1.5)	Average size for 480p video is 250 MB per hour in YouTube [85]
VoIP data (G726 codec [81])	Duration	900 s (+/−300)	Skype audio only UL 42–47 kbps, DL 42–47 kbps [82]
	Rate	45 kbps UL & DL	Average Skype call duration is ∼20 minutes [83]

Table 10
Types of devices and associated battery consumption.

Device type	Screen type	Battery capacity (mAh)	Macro cell battery consumption	Femto cell battery consumption
High capabilities device (Based on Nexus 5)	LCD 1080*1920, 4.95inch	2300	145,27*t(talk) + 483,19*t(web) + 377,04*t(video) + 7,66*t(idle)	0.714 * Macro cell consumption
Medium capabilities device (Based on Samsung S3)	Super AMOLED 720*1980, 4.8inch	2100	215,38*t(talk) + 325,58*t(web) + 222,22*t(video) + 2,65*t(idle)	
Low capabilities device (Based on Nokia E66)	TFT 240*320, 2.4inch	1000	133,33*t(talk) + 230,77*t(web) + 312,5*t(video) + 2,97*t(idle)	

imum of 2 GB (the initial value is randomly generated between 100–2000 MB), and Gold users have no threshold.

When data services are used, the available data that a user has according to his subscription are reduced. Recall that in real life, users tend to reduce their activities (i.e. data usage and session duration) when their data availability becomes low. In order to replicate this behavior, we assume that when bronze users consume 80% of available data they reduce their activities to 10% of their normal habits and corresponding session duration to 70%. Similarly, silver users reduce their data intensive activities to 50% when they consume 80% of their available data and cut their duration by 50%.

We consider 3 different device classes, namely high, medium and low capabilities terminals, which affect the total battery capacity and batter reduction of a UE. The battery consumption formula is based on the battery's maximum capacity and the battery's consumption of each service on every device. In the case of smaller (Femto) cells, we consider that the consumption is proportional to the Macro cell's consumption [86–88] due to the fact that less transmission power is needed on the UE side. Throughout the simulation we consider that users do not change devices.

Battery status is initialized similarly to charging status (i.e. uniform distribution, but taking into account maximum capacity due to different devices). Furthermore, we consider 3 distinct battery levels, namely *High, Medium* and *Low*, each having an impact on the user behavior. Initial battery state is randomly distributed between 20% and 80% of maximum capacity. A device has High battery status when more than 35% of the total battery capacity is available. A high battery status has no effect in the rate or duration of the services used throughout the scenario. A Medium battery status is assigned to devices operating between 10% and 35% of total battery capacity. When a device's battery status drops to Medium, the user consumes 50% of the calls/services he would normally consume (i.e. in High battery status) and their duration is cut by 50%. Finally, a low battery status is assigned to devices operating below 10% of total battery capacity. When a device's battery status drops to Low, the user only uses 10% of calls/services he would normally consume (i.e. in High battery status) and their duration is reduced by 70% (duration reduction does not apply for web pages). Table 10 below describes in detail the characteristics of the three device classes and the battery models.

3.6.2. Experimentation methodology

Using this scenario and assumptions we generate a number of datasets upon which CEPE was evaluated. During simulation time we monitor the status of the user, device, network and service and store it into a log file. Each record of the dataset contains the time of the observation, the UE's International Mobile Subscriber Identity (IMSI) and all of the information we can collect related to UE's, services' and network's state at the specific time of the observation. An excerpt of this information is depicted in the following figures (Figs. 4–6). Note that in the device dynamic measurements (Fig. 5) we have also considered user mobility as collected in the simulation environment in a simplified way. In a real implementation of CEPE this information can be collected by reported information from the UE based on GPS data, or network collected information like the recorded positions of network attachment or the handover rate.

Data are post-processed in order to add labeling information; every instance is attributed a label based on the scenario assumptions of the previous paragraph. We consider this labeling information as the ground truth, i.e. the correct labels that our model should identify. Moreover, we map continuous variables like income and age to nominal values and string data to integer values (e.g. an education level 'College' is mapped to 2). This way we manage to map all variable values to real numbers and thus map any instance as a high dimensional point residing in R^n, where n is the number of variables describing each observation.

The final dataset is provided to CEPE for training. As soon as the knowledge base is built we rerun the same experiment using the derived rule-set and the learnt model. The second time, NS-3 uses the CEPE model in order to i. classify a user according to his behavior and ii. identify the best set of actions to apply given the results of i.

Evaluation is performed along two axes; at first we assess the classification capability of CEPE itself and then we assess its effect on network conditions. Both supervised and unsupervised CEPE versions are evaluated with respect to their ability to build a model

imsi	name	surname	age	gender	education	income	cpu	cores	os	screen_width	screen_height
1	Maria	Taylor	36	F	College	33220	1.3Ghz	dual	iOS	240	320
2	Nick	Lukas	19	M	not graduated Highschool	24712	1.3Ghz	dual	iOS	240	320
3	George	Brown	22	M	Higher than College	22394	1.7Ghz	dual	Android OS	720	1280
4	George	Smith	29	M	Higher than College	24916	1.6Ghz	quad	Android OS	1080	1920
5	Lusille	Brown	54	F	College	47098	1.6Ghz	quad	Android OS	1080	1920
6	George	Tremblay	58	M	Higher than College	56884	1.3Ghz	dual	iOS	240	320
7	Nick	Smith	78	M	College	16722	1.6Ghz	quad	Android OS	1080	1920
8	Constantine	Smith	50	M	not graduated Highschool	18461	Unknown	Unknown	Symbian	240	320
9	John	Smith	51	M	Highschool	23208	Unknown	Unknown	Symbian	240	320
10	Luisse	Skywalker	43	F	Higher than College	15122	1.3Ghz	dual	iOS	240	320
11	George	Skywalker	44	M	not graduated Highschool	23320	1.7Ghz	dual	Android OS	720	1280
12	Nick	Skywalker	28	M	Higher than College	181582	1.3Ghz	dual	iOS	240	320
13	George	Wong	30	M	College	77579	1.3Ghz	dual	iOS	240	320
14	Martha	Brown	21	F	Higher than College	5488	1.7Ghz	dual	Android OS	720	1280
15	Maria	Brown	50	F	College	30939	1.6Ghz	quad	Android OS	1080	1920

Fig. 4. Sample of the user and device static characteristics.

time	imsi	x	y	z	velocity (x:y:z)	txpower	ultx	sum_ultx	dlrx	sum_dlrx	ul_delay	dl_delay	lost_packets	associated_cellid	rsrp	rsrq
375	5	18.1683	25.2491	9.08511	00:00:00	20	0	4210	18	17364	0	0.111069	0	2	-79.0004	-6.0066
375	21	52.2987	68.3325	11.1514	-1.95099:-1.56354:0	20	25	11359	26	20696	0.0123813	0.00714496	0	1	-77.2772	-5.48997
375	22	69.5819	24.7065	4.79334	0.770973:1.97211:0	20	0	10944	0	25848	0	0	0	8	-78.8653	-8.30015
375	23	73.6172	76.5321	4.62715	-2.20642:1.29282:0	20	25	10720	25	24558	0.0122713	0.00334262	0	1	-77.2884	-6.45531
375	24	45.9436	7.94052	4.61572	1.0997:-2.01332:0	20	25	9964	25	23387	0.0126559	0.00372722	0	6	-66.9001	-4.2853
375	25	38.2861	90.6926	11.9442	-2.54065:0.860013:0	20	0	13824	19	25391	0	0.554631	0	1	-76.0733	-6.23794
375	26	30.7951	40.5527	12.9944	2.62772:-0.491851:0	20	0	12020	18	25496	0	0.552982	0	1	-78.2644	-5.32413
375	27	70.3948	28.1009	6.23434	3.48727:1.06985:0	20	25	11327	15	25341	0.0124184	0.109823	0	8	-76.8916	-8.68234
375	28	140.162	49.6648	11.8284	-0.034281:3.55078:0	20	25	11309	25	23303	0.012905	0.0244163	0	1	-76.3161	-6.0656
375	29	100.556	51.5388	6.39348	0.930197:-2.26329:0	20	0	10673	0	22124	0	0	0	11	-78.6731	-5.80526
375	30	171.34	26.2544	2.14341	3.83806:-0.799411:0	20	0	11034	0	23307	0	0	0	16	-75.897	-7.50523
375	31	170.5	87.0532	5.88935	-1.42696:-3.05206:0	20	0	6393	0	20687	0	0	0	1	-66.8345	-4.11972
375	32	14.1901	78.9689	7.97035	-0.515784:2.77083:0	20	25	9791	26	20536	0.0124892	0.00613741	0	1	-76.4907	-5.66534
375	33	186.025	39.6246	9.94747	-0.446031:-2.35723:0	20	0	8065	0	17411	0	0	0	1	-80.7119	-6.22565
375	34	194.926	67.5309	3.0497	0.546728:-2.56947:0	20	0	8337	11	21214	0	0.154213	0	21	-69.1081	-4.34456
375	35	83.7675	10.6315	7.6823	2.71138:0.0945869:0	20	0	9740	0	22202	0	0	0	8	-71.2933	-6.08535
375	36	144.869	19.5153	12.0044	2.99819:1.88492:0	20	0	11859	0	22371	0	0	0	16	-75.2858	-7.65617
375	37	91.7005	90.1601	11.3707	-3.11177:1.98614:0	20	0	10609	0	19951	0	0	0	1	-77.1156	-7.36175
375	38	33.7284	98.4918	0.251241	0.824956:-3.56771:0	20	0	7209	0	23529	0	0	0	5	-70.8778	-5.85312
375	39	50.1913	89.3476	14.7854	-1.99062:1.13374:0	20	0	9910	24	20846	0	0.488858	0	1	-76.6427	-6.00061
375	40	67.6137	3.07486	6.3801	-3.118:1.33938:0	20	0	7128	0	20673	0	0	0	1	-70.8111	-5.38928

Fig. 5. Sample of user and device dynamic measurements.

cellid	cell_type	x2	y2	z2	txpower5	resource_blocks	dltx	sum_dltx	ulrx	sum_ulrx	lost_packets6	#ues	connected_imsis
2	Femto_Cell	10	10	1	10	15	25	20851	0	3891	7	1	5
1	Macro_Cell	0	250	6	35	50	25	32665	25	10371	-1	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
8	Femto_Cell	70	10	1	10	15	0	32321	0	9925	0	4	8:22:40:27
1	Macro_Cell	0	250	6	35	50	25	32994	25	9224	0	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
6	Femto_Cell	50	10	1	10	15	25	31976	25	8851	0	2	13:24
1	Macro_Cell	0	250	6	35	50	25	35694	0	12306	6	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
1	Macro_Cell	0	250	6	35	50	25	34263	0	10521	7	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
8	Femto_Cell	70	10	1	10	15	25	34381	25	10382	10	4	8:22:40:27
1	Macro_Cell	0	250	6	35	50	25	33212	25	10171	0	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
11	Femto_Cell	90	91	1	10	15	0	32424	0	8986	0	0	
16	Femto_Cell	150	10	1	10	15	0	33718	0	9356	0	2	07:36
1	Macro_Cell	0	250	6	35	50	0	30702	0	5787	0	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
1	Macro_Cell	0	250	6	35	50	25	32005	25	9077	-1	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
1	Macro_Cell	0	250	6	35	50	0	29326	0	7448	0	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
21	Femto_Cell	190	91	1	10	15	3	32288	0	7849	-1	1	34
8	Femto_Cell	70	10	1	10	15	0	33929	0	9604	0	4	8:22:40:27
16	Femto_Cell	150	10	1	10	15	0	33329	0	10525	0	2	07:36
1	Macro_Cell	0	250	6	35	50	0	32004	0	8971	0	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
5	Femto_Cell	30	91	1	10	15	0	31780	0	6683	0	2	09:38
1	Macro_Cell	0	250	6	35	50	25	31093	0	8654	1	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28
1	Macro_Cell	0	250	6	35	50	0	30484	0	6181	0	16	10:15:16:17:20:6:26:25:39:23:37:32:33:29:21:28

Fig. 6. Sample of the network related measurements for the same time period.

that identifies new observations. In the unsupervised CEPE case we assess

i. the ability of Spectral Clustering to build a model that is close to the original labeling scheme
ii. and the ability of ID3 to properly model it and classify new instances with it.

The evaluation of the supervised case is simpler since we only assess the ability of CEPE to build a model which is close to real data.

In order to assess the quality of the clustering and classification algorithms we employ the $F-measure = 2*\frac{Precision*Recall}{Precision+Recall}$, where $Precision = \frac{True\ Positives}{True\ Positives+False\ Positives}$ and $call = \frac{True\ Positives}{True\ Positives+False\ Negatives}$. F-measure is a binary classification metric (i.e. 1–0 classification) however it can be extended for the multiclass case. In extending a binary metric to multiclass or multilabel problems, the data is treated as a collection of binary problems, one for each class. So, we calculate the F-measure per class and then average obtained values across the set of classes.

Due to the inherent randomness of k-Means we validate our results by applying a 10-fold cross validation procedure. During this process we break the dataset into 10 equisized bins; we use 9 bins for training the model and 1 for testing it. We repeat the process 10 times each time using a different testing bin, so as to make sure

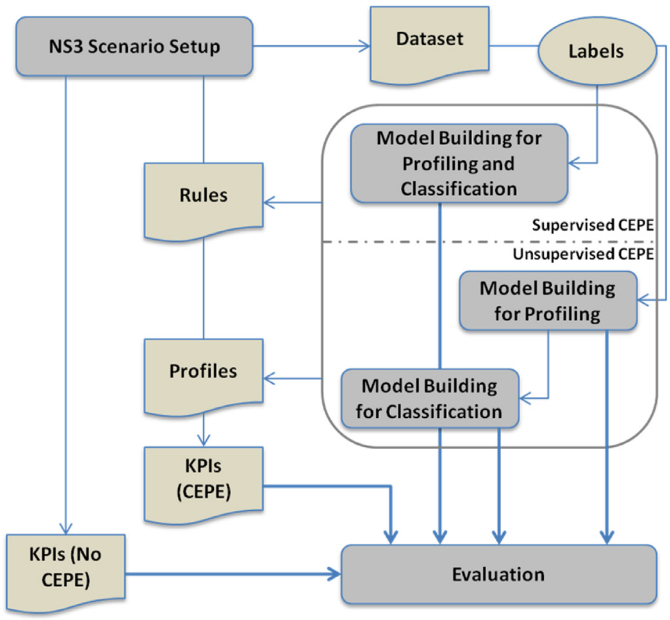

Fig. 7. Experimentation methodology.

that all bins have been used for both training and testing. In every iteration we compute the F-measure as describe in the previous paragraph and report the average value obtained throughout the 10 iterations.

We retain the best performing model and proceed along the second evaluation axis that focuses on the effect of CEPE rules and profiles on network conditions. We quantify mainly two KPIs, namely: *Uplink (UL)/Downlink (DL) Throughput* (i.e. the total throughput for the uplink and downlink respectively, for the entire simulation) and *Number of Handovers* (i.e. the total handovers performed during the entire simulation). The latter, i.e. *Number of Handovers*, is directly linked to QoS: the more handovers are realized, the higher the degradation of the respective on-going service performance, due to the handover signaling overhead increase, data lost during the handover duration, re-transmission timeouts, etc. The overall experimentation methodology is depicted in Fig. 7.

Although there are numerous schemes for RAT selection that -in their majority- improve the network-related KPIs, when comparing to the baseline LTE systems it was not straightforward to select the most representative ones so as to perform a comparative assessment of the CEPE-induced added value, since different proposed mechanisms focus on different KPIs. In addition, diverse approaches are followed (utility functions, neural networks, MADM, etc.); last but not least different simulation set-ups and tools are being chosen by each researcher thus making a direct comparison even less straightforward. So we compare the aforementioned KPIs before and after the application of CEPE.

We employed two simulation models: Low Background Traffic, where during experiments current traffic was augmented by a +5% UL and DL and Medium Background Traffic, where current traffic was augmented by +20% UL and DL. The duration of each simulation equals 20 real-life minutes and we logged measurements every second. We considered 40 users, using the discussed service types (5 types), mobility patterns (3 patterns), device types (3 types) and charging categories (3 categories). Every scenario was executed 5 times thus all reported values in the rest of the section comprise the mean results of these executions.

4. Results

The application of the supervised CEPE resulted in a model, which attains an F-measure of 0.984 (e.g. the derived rules from ID3 can correctly classify 98.4% of observations). The background traffic did not affect the classification ability of the models. The unsupervised CEPE was more laborious since it entailed the initial application of Spectral Clustering. Specifically, we used a fully connected graph and the Gaussian kernel with $\sigma = 3$.[3] Spectral clustering managed to attain an F-measure of 0.90 (e.g. identified the correct number of clusters and correctly classified 90% of the instances). The application of the decision trees on the labeling scheme of the spectral clustering produced an F-measure of 0.95. Table 11 provides the aforementioned results both holistically but also per entity (e.g. device entities only). Since it is easier to build a good model for Devices but it gets more difficult when the device classes are combined with the various mobility types and services we decided to present the dynamic information (i.e. mobility data) separately.[4]

The experiments showcase a clear superiority of the supervised case, constantly exhibiting a better F-measure. In fact, when the unsupervised CEPE was evaluated against the ground truth labeling scheme, the F-measure was marginally equal to 0.90 giving the supervised case a clear precedence of almost 10%.

It is also worth noting that in the supervised case, classification results and associated rules are derived faster. The latter is due to the complexity of the employed approach; spectral clustering necessitates $O(|\mathbf{X}|^2)$ memory and $O(|\mathbf{X}|^3)$ processing time contrary to $O(|\mathbf{X}|)$ memory and $O((n|\mathbf{X}|)^2)$ time required by decision trees -n is the number of features.

This evaluation also indirectly assesses the capability of CEPE to identify behavioral changes and adapt (i.e. identify that a subscriber suddenly changes mobility pattern); this is captured in the overall results. Finally, the lower quality was anticipated due to the unsupervised nature of the algorithm. Thus, in the remainder of the section we focus on supervised CEPE.

Using the methodology of Section 3 and exploiting the derived classes-graph we extracted a number of rules. We used the semi supervised approach where we extracted the full rule-set from the graph, ranked it according to the sum of weights and selected the rules which we deemed more suitable for ameliorating the KPIs. The selection we performed was based on the following assumptions / considerations:

- Gold users should have the highest possible quality of experience.
- High mobility users shall be served by macro-cells to minimize the number of handovers.
- Calls of higher duration by moving users shall be served by macro-cells since the probability of a handover during the lifetime of the call is higher than for short calls.
- Voice Services shall be served by a second/third generation technology so that resources in 4G systems are allocated to high data users.
- Low-end devices that cannot support advanced services should be served by a Second or Third Generation technology (e.g., GSM).
- Low Battery level devices shall use short-range wireless technology.[5]

[3] The derived affinity matrix offered the best discriminative depiction of the underlying data clusters (i.e. clusters were better separated and clearer compared to other configurations).

[4] Note the 5%-10% difference in the exhibited F-measure between Device and Mobility; combining these together would indirectly conceal this.

[5] Note here that short-range wireless technology does not always imply lower power consumption, since this also involves numerous other parameters related to

Table 11
Unsupervised and supervised CEPE results -knowledge discovery capability assessment w.r.t F-measure.

CEPE KDD evaluation		*Unsupervised*		*Supervised*
		Spectral clustering	*Decision trees*	*Decision trees*
Low background traffic	*Device*	0.97	1	1
	Mobility	0.89	0.951	0.981
	User	0.95	1	1
	Service	0.91	0.934	0.973
	All	**0.90**	**0.95**	**0.984**
Medium background traffic	*Device*	0.97	1	1
	Mobility	0.89	0.951	0.981
	User	0.95	1	1
	Service	0.91	0.934	0.973
	All	**0.90**	**0.95**	**0.984**

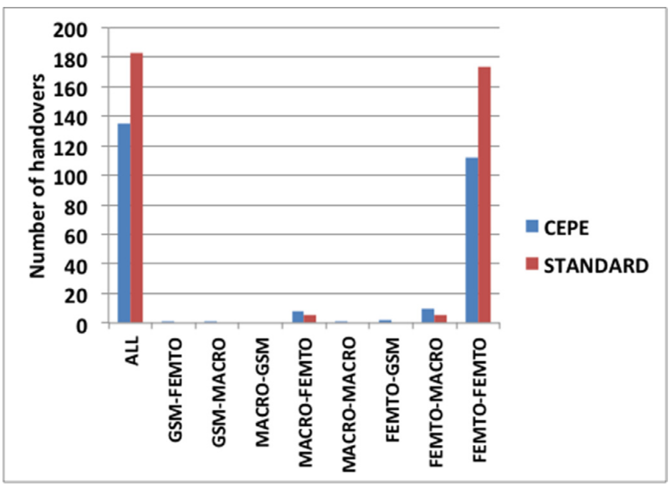

Fig. 8. Number of handovers per RAT type - low background traffic experiments.

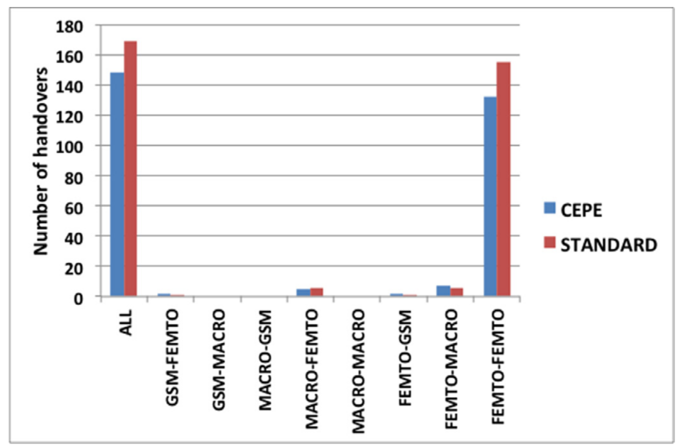

Fig. 9. Number of handovers per RAT type - medium background traffic experiments.

These rules are used in conjunction with the A2A4 handover algorithm. Each rule selects the most appropriate RAT (e.g., FemtoCell), and A2A4 undertakes the selection of the best FemtoCell where a UE should be handed over, based on RSRQ.

4.1. Performance evaluation per RAT

Our first evaluation depicts the results in relation to the RAT of the UE. Essentially, for every employed KPI we attempt to quantify it on a holistic level (e.g. how the packet loss was affected by CEPE) as well as on RAT level (how the packet loss in GSM was affected by CEPE). Towards this end we provide Figs. 8–11.

The four initial figures illustrate the overall number of handovers per RAT type, as well as the experienced throughput per RAT, both for medium and low background traffic. The graphs show that the application of CEPE reduces the overall number of handovers in both cases; in fact, the realized handovers are minimized by 15 – 20%. Looking closer, we observe that the femto-femto handover type dominates the overall picture. The reasoning for this fact is primarily related to the Ultra Dense Environment of the simulation topology that has been selected, and not the specific policy

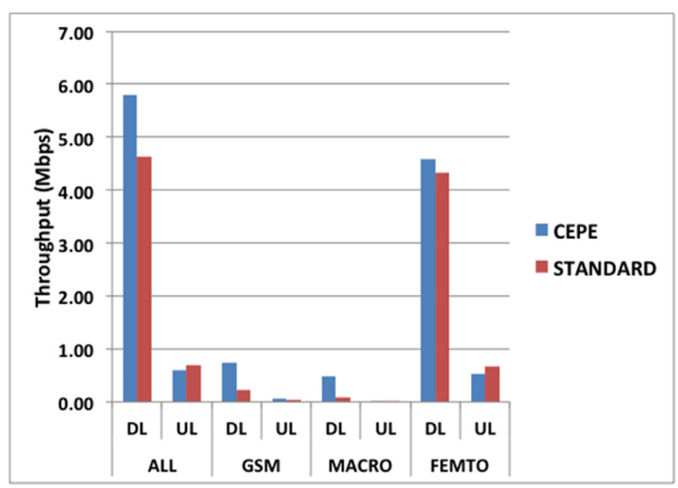

Fig. 10. Experienced throughput per RAT - low background traffic experiments.

rules that were applied. There is a big number of femto cells inside the interior of the shopping mall, very close to which the UEs are moving. This results to high RSRP/RSRQ values, boosting as a result the femtos' selection.

Despite this fact, however, CEPE seems to considerably ameliorate the ping-pong effects and the often handovers – even in the very frequent event of femto to femto handover-; this is on the one hand due to the policy rule that high mobility users are never placed in femto cells (thus these users never participate in FEMTO-FEMTO handovers); on the other hand, the rule, which places the

MAC protocol specification, Rx and Tx radio frequency components (RF), cell bandwidth, discontinuous reception (DRX), etc. In our experimental configuration however and due to several NS3 simulator's limitations, we decided to apply some specific battery consumption models (e.g. Table 9, [86–88]) which primarily assume that the UE needs to boost its transmission power towards interference mitigation in cases of wide-range access points. According to our assumptions, such requirement becomes even more impelling in an ultra dense environment with multiple co-existing femto cells in a limited geographical area.

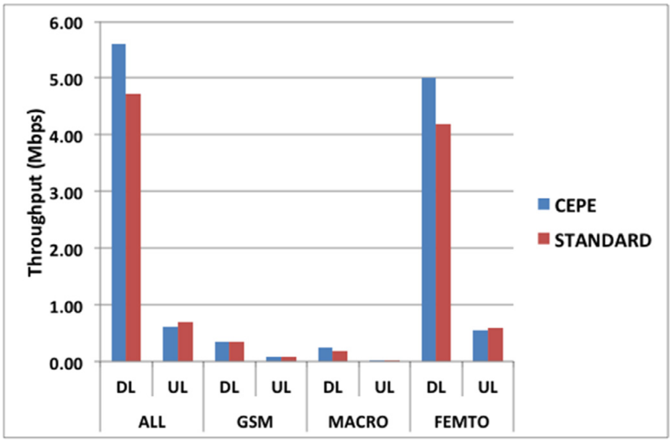

Fig. 11. Experienced throughput per RAT - medium background traffic experiments.

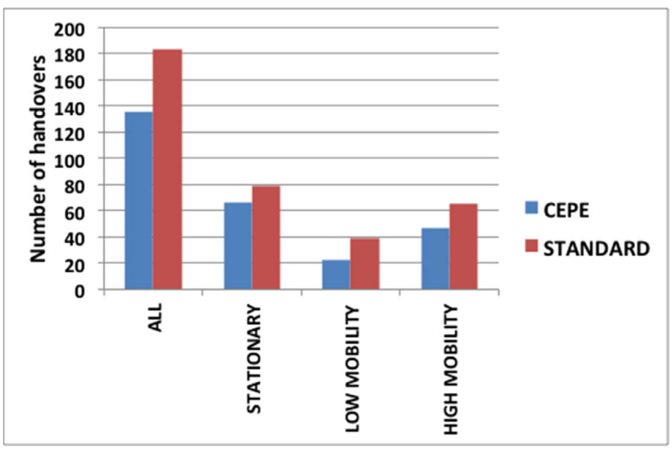

Fig. 12. Number of handovers per mobility type- low background traffic experiments.

Voice Service UEs in the macro cells also decreases the possibility for UEs to need to realize a handover between adjacent FEMTO cells.

It is worth mentioning the fact that we observed additional advantages in the ul/dl packet loss and delay (although we are not employing them as assessment KPIs). In the CEPE-enabled case the packet loss for the downlink, which accounts for ~90% of the traffic, was decreased by more than 4%. Moreover in the CEPE-enabled handover scheme the delay for the downlink was also decreased by almost 10%. Similarly to the throughput results' analysis, downlink delay was decreased for the LTE macro cells and femto cells since CEPE allocated each Service to the appropriate RAT considering QoS requirements and mobility type. The downlink delay was increased for the GSM BS (i.e. Voice calls) due to the increased distance and the different configuration of GSM comparing to other LTE technologies; still however, it is in the acceptable limits (i.e., below 200 ms).

4.2. Performance evaluation per mobility type

The second evaluation perspective illustrates the simulation outcomes in relation to the type of UEs mobility. Similarly to the previous paragraph, we provide all the results in Figs. 12–15, presenting the overall number of realized handovers and experienced throughput, both for low, as well as for medium traffic experiments.

The CEPE-enabled network manages to decrease the overall number of handovers in both scenarios and for all mobility type cases. Only the stationary users in the medium traffic scenario experience the same number of handovers. The minimization of the overall handovers is achieved by the rules tendency to place each user to the optimal RAT according to the profile, which has been extracted and avoid consecutive handovers, based only on the RSRQ metric.

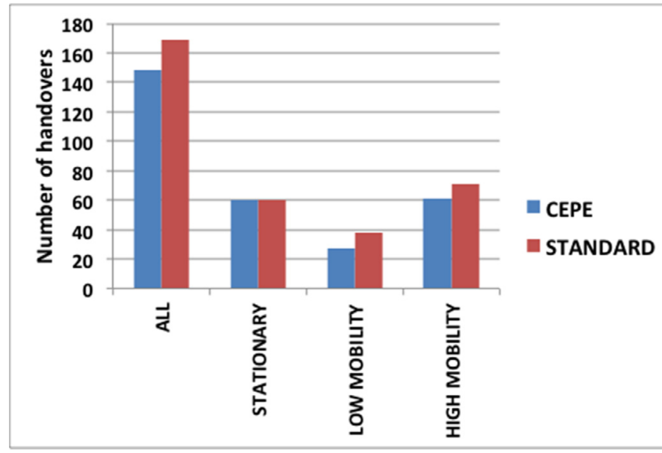

Fig. 13. Number of handovers per mobility type - medium background traffic experiments.

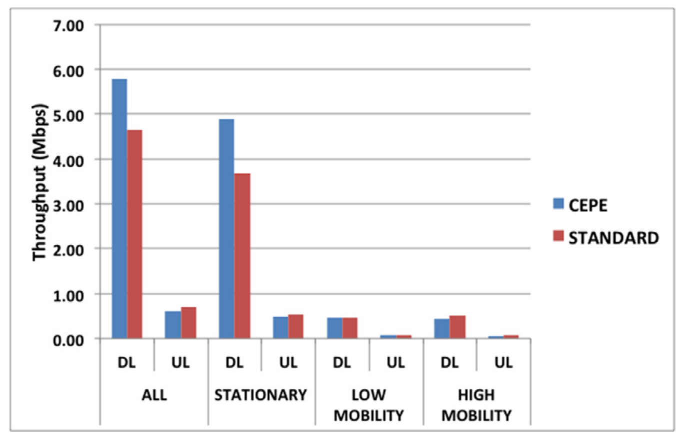

Fig. 14. Experienced throughput per mobility type- low background traffic experiments.

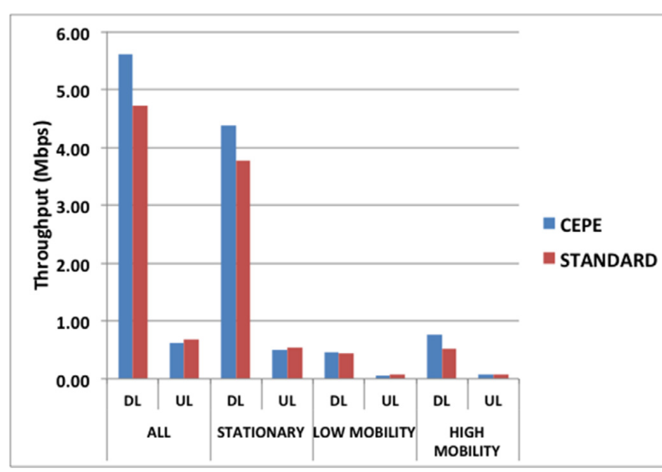

Fig. 15. Experienced throughput per mobility type - medium background traffic experiments.

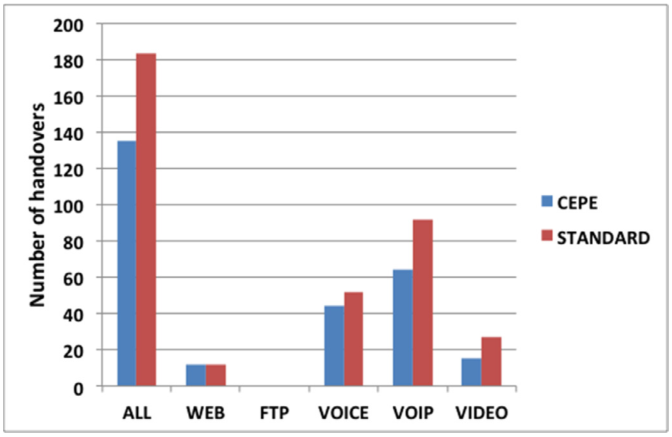

Fig. 16. Number of handovers per service type- low background traffic experiments.

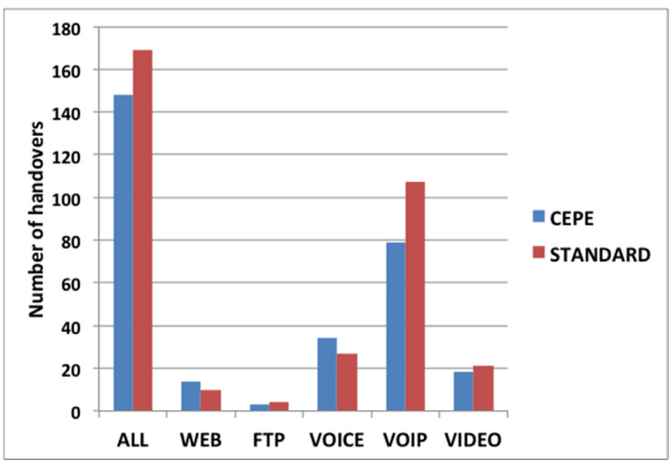

Fig. 17. Number of handovers per service type - medium background traffic experiments.

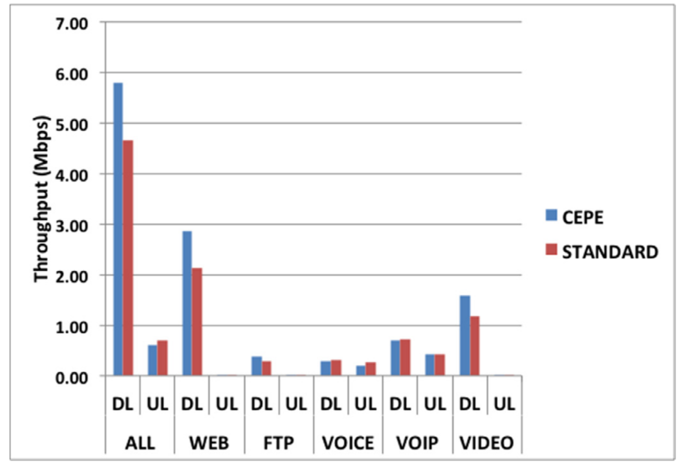

Fig. 18. Experienced throughput per service type- low background traffic experiments.

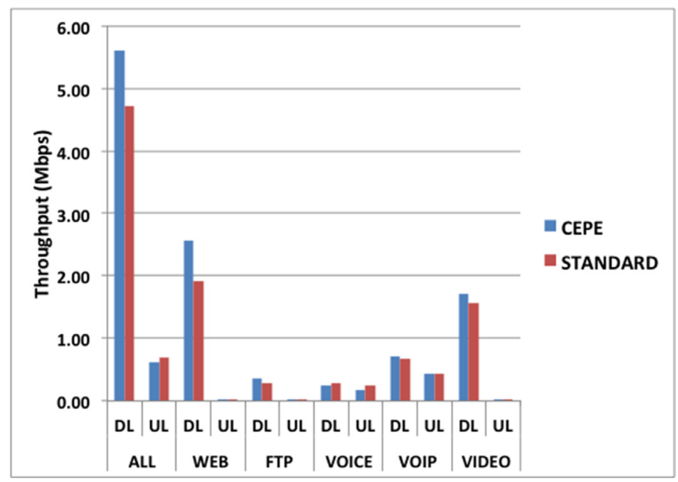

Fig. 19. Experienced throughput per service type - medium background traffic experiments.

The experienced throughput in the CEPE-enabled handover scheme case is also improved in both traffic scenarios. The most significant gain is observed in the case of stationary users; in the low background traffic scenario the average gain for CEPE is equal to 33%, while in the medium background traffic it is slightly over 19%. Note that the CEPE-enabled scheme shows similar performance to the standard scheme in almost all other cases. In addition, we observe that the mobility is inversely correlated with the throughput gain; the lower the mobility the higher the throughput gains. This is mainly associated with the rule that switches all stationary users with high throughput Services (e.g., Web or FTP) to femto cells. In the case of uplink, a slight decrease (i.e., 2–6%, depending on the case) is observed in all mobility types, mainly due to the interference. The latter is the result of the fact that the number of users that are associated to a macro cell in the CEPE experiment is four times higher when comparing to the standard.

4.3. Performance evaluation per service type

During the simulations, diverse types of services were deployed in the UEs. The possible service types were discussed in Table 9. Similarly to the previous paragraphs, we provide all the results in Figs. 16–19.

By studying the graphs one will notice that voice traffic is on FEMTO while it could have been served by GSM. The latter is due to the fact that our rules operate in conjunction with the A2A4 RSRQ handover algorithm. This means that the first prerequisite for realizing a handover is that the candidate target cell satisfies the A2A4 algorithm requirements and thresholds; the described rules apply at the second step. During our simulations, when the A2A4 RSRQ requirements were met, several voice traffic handovers took place towards a GSM. In all the other cases (i.e., when only FEMTO cells are satisfying the RSRQ thresholds), a voice traffic handover to an optimal FEMTO cell is realized as well.

The first KPI, which is illustrated in the figures, is the overall number of handovers that took place per Service type. The latter is decreased for all service types when deploying the CEPE-enabled handover scheme, apart from the Web type of service case, in which the CEPE scheme realizes equal number of handovers with the standard scheme. The CEPE-enabled handover mechanism shows an overall enhancement in both traffic scenarios in the downlink case, while regarding the uplink one, the performance of the two schemes is almost identical with a minor decrease in the low and medium traffic cases. It should be noted that the higher gain for the CEPE-enabled scheme is observed in the Web and the Video service types. The observed amelioration in the throughput KPI per Service is due to the enhanced allocation of the users to the respective RATs based on the Services properties and requirements. Note that the Voice/VoIP are practically the same whereas for the more demanding ones we have more benefits.

5. Conclusion

In the context of this paper we proposed the derivation and dynamic update of a knowledge base and a rule-set catering for the optimization of network operation by means of an overarching knowledge discovery framework, namely CEPE. CEPE collects and subsequently processes the information monitored through the lifetime of a network and opens new horizons on how to create automatically the profile of users and use it to enhance the performance of network control functions and subsequently the overall performance of the network. Through a comprehensive literature review we demonstrated the need for such a framework. Thereinafter we provided an instantiation of CEPE using well known data mining and machine learning solutions. By means of an extensive experimentation effort, we assessed the validity and viability of our proposal in a close-to-real-world environment.

Our future research efforts will focus on advancing CEPE; we intend to perform further experiments in order to assess the applicability of the unsupervised CEPE flavor, the automatic rule extraction mechanism and the associated network requirements. In parallel, we will also extend the scenarios to accommodate CAC and cell-(re)selection aspects. In the same context, we will investigate the necessary implementation technologies (e.g. real-time processing software, big data solutions etc), architectures and their mapping on network elements so as to render CEPE as a possible real-life solution. Lastly, we will study the theoretic properties of the framework, including a proof of convergence for the rules evaluation approach and assess its application on related network problems like traffic engineering.

Acknowledgment

The authors would like thank Mr. Manolis Kiagias for his contribution in this research effort.

References

[1] Cisco Visual Networking Index, 2017. Global mobile data traffic forecast update, 2015–2020 White Paper[accessed January [accessed January 2017] http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.
[2] A. Hakiri, P. Berthou, Leveraging SDN for the 5G networks: trends, prospects and challenges, Book Chapter in "Software Defined Mobile Networks (SDMN): Beyond LTE Network Architecture, John Wiley & Sons, Ltd., 2015.
[3] 3GPP TS 36.304, “User equipment (UE) procedures in idle mode (Release 12)”, Vol. 12.6.0, September 2015.
[4] F. Tesema, A. Awada, I. Viering, M. Simsek, G. Fettweis, Fast cell select for mobility robustness in intra-frequency 5G ultra dense networks, 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), September 2016.
[5] S. Kumar, A. Kumar, K. Detroja, Fuzzy logic based cell selection framework for downlink uplink decoupled cellular networks, 22nd National Conference on Communication (NCC), 2016 March.
[6] T. Qu, D. Xiao, D. Yang, W. Jin, Y. He., "Cell selection analysis in outdoor heterogeneous networks", 3rd IEEE International Conference on Advanced Computer Theory and Engineering (ICACTE), 2010.
[7] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, J. Tingfang, Cell association and interference coordination in heterogeneous LTE-a cellular networks, Selected Areas in Communications, IEEE Journal, 2010.
[8] Z. Becvar, P. Mach, "Performance of fast cell selection in two-tier OFDMA networks with small cells", IFIP Wireless Days (WD), 2012
[9] M. Bembe, K. Jeongchan, T. Olwal, H. Youngnam, Available bandwidth-aware cell selection for expanded regions of small cells adopting ABS, International Conference on Information and Communication Technology Convergence, 2013.
[10] H.S. Dhillon, J.G. Andrews, Downlink rate distribution in hetnets cellular networks under generalized cell selection, IEEE Wireless Commun. Lett. 3 (1) (June 2013).
[11] R. Thakur, S. Mishra, C. Siva Ram Murthy, "A load-conscious cell selection scheme for femto-assisted cellular networks", IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2013.
[12] J. Oh, Y. Han, Cell selection for range expansion with almost blank subframe in heterogeneous networks, IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2012.
[13] J. Sangiamwong, Y. Saito, N. Miki, T. Abe, S. Nagata, Y. Okumura, Investigation on cell selection methods associated with inter-cell interference coordination in heterogeneous networks for LTE-advanced downlink, in: Proceedings of European Wireless Conference, 2011.
[14] M. Feng, X. She, L. Chen, and Y. Kishiyama, “Enhanced dynamic cell selection with muting scheme for DL CoMP in LTE-A”, Proceedings of the IEEE Conference on Computer Communications (INFOCOM), 2010.
[15] A.Daeinabi, K. Sandrasegaran, X. Zhu, Performance evaluation of cell selection techniques for picocells in LTE-advanced networks, 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2013.
[16] J. Wannstrom, K. Mallinson, Heterogeneous networks in LTE, 2015. 3GPP, available from: November http://www.3gpp.org/technologies/keywords-acronyms/1576-hetnet.
[17] L. Gao, X. Wang, G. Sun and Y. Xu, “A game approach for cell selection and resource allocation in heterogeneous wireless networks”, IEEE International Conference on Sensing, Communication and Networking (SECON), 2011.
[18] S. Kim, Y. Lee, “Adaptive MIMO mode and fast cell selection with interference avoidance in multi-cell environments,” 5th International Conference on Wireless and Mobile Communications (ICWMC), 2009.
[19] H.-S. Jo, Y.J. Sang, P. Xia, J.G. Andrews, Heterogeneous cellular networks with flexible cell association: a comprehensive downlink SINR analysis, IEEE Trans. Wireless Commun. 11 (10) (July 2011).
[20] Patent EP 2529581 A1, “Cell selection and reselection in a telecommunication network”.
[21] Patent EP 2647246 A1, “Cell re-selection using a ranking algorithm”.
[22] Patent EP 2292050 A2, “Cell selection and reselection in deployments with home NodeBs”.
[23] Patent EP 2458919 B1, “Cell re-selection in a cellular telecommunications network”.
[24] Patent EP 2387279 A1, “Cell (re)selection in a heterogeneous mobile communication system”.
[25] M. Ghaderi, R. Boutaba, Call admission control in mobile cellular networks: a comprehensive survey, Wireless Commun. Mob. Comput. 6 (1) (2006) February.
[26] M. Ahmed, “Call admission control in wireless networks: a comprehensive survey”, IEEE Commun. Surv. Tut., Vol. 7(1), 2005.
[27] D. Niyato, E. Hossain, “Call admission control for QoS provisioning in 4G wireless networks: issues and approaches”, IEEE Netw., Vol. 19 (5), 2005.
[28] L. Yun Yi, A.R. Mohd Shariff, K. Mohd Zaini, User behavior based call admission control for traffic steering in Wi-Fi/cellular networks, Adv. Comput. Commun. Eng. Technol. 362 (2015) 407–417 December.
[29] M. Jain, R. Mittal, Adaptive call admission control and resource allocation in multi server wireless/cellular network, J. Ind. Eng. Int. 12 (March 1) (2016) 71–80 Issue.
[30] H.S. Ramesh Babu, Gowrishankar, P.S. Satyanarayana, Call admission control mechanism for optimal QoS in next generation wireless networks, International Conference on Intelligent Systems, Modelling and Simulation, 2010.
[31] O.E. Falowo, H. Anthony Chan, Joint call admission control algorithm for fair radio resource allocation in heterogeneous wireless networks supporting heterogeneous mobile terminals, IEEE Consumer Communications & Networking Conference, 2010.
[32] J. Sabari Ganesh, P.T.V. Bhuvaneswari, “Enhanced call admission control for WiMAX networks”, IEEE-International Conference on Recent Trends in Information Technology (ICRTIT), 2011.
[33] S.-Q. Lee, R.B. Han, N.-H. Park, Call admission control for hybrid access mode femtocell system, 4th IEEE International Workshop on Selected Topics in Mobile and Wireless Computing, 2011.
[34] S. Sha, R. Halliwell, Performance modelling and analysis of dynamic class-based call admission control algorithm using fuzzy logic for heterogeneous wireless networks, International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11, 2011.
[35] C. Olariu, J. Fitzpatrick, P. Perry, L. Murphy, A QoS based call admission control and resource allocation mechanism for LTE femtocell deployment, IEEE Consumer Communications and Networking Conference - Wireless Consumer Communication and Networking, 2012.
[36] S.A. Sharna and M. Murshed, “Impact on vertical handoff decision algorithm by the network call admission control policy in heterogeneous wireless networks”, IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC), 2012.
[37] Y. Kim, H. Ko, S. Pack, W. Lee, X(.Sherman). Shen, Mobility-aware call admission control algorithm with handoff queue in mobile hotspots, IEEE Trans. Vehic. Technol. 62 (8) (October 2013).
[38] Patent EP 2580929 A1, “Admission control for shared LTE network”.
[39] Patent EP 1805938 A2, “Methods and systems for measurement-based call admission control in a media gateway”.
[40] Patent EP 1661334 B1, “Call admission control system and method for interpreting signaling messages and controlling traffic load in internet protocol differentiated services networks”.
[41] Patent EP 1641232 A1, “Call admission control in a VoIP network”.
[42] Patent EP 2249544 A1, “Call admission control device and call admission control method”.
[43] Patent EP 2317811 A1, “Call admission control device and call admission control method”.
[44] Patent EP 1471764 B1, “Call-admission controller and method of call-admission control”.

[45] Patent EP 1796413 B1, "Call admission control device and call admission control method in a wireless communication system".
[46] Patent EP 1671512 B1, "Adaptive call admission and policing control for a communication link with limited bandwidth".
[47] Patent EP 1715638 B1, "Method and apparatus for quality-of-service-based admission control".
[48] 3GPP 23.829 – a01 V10.0.1, "Local IP access and selected IP traffic offload (LIPA-SIPTO)", Release 10, October 2011.
[49] D. Xenakis, N. Passas, L. Merakos, C. Verikoukis, "Mobility management for femtocells in LTE-advanced: key aspects and survey of handover decision algorithms", IEEE Commun. Surv. Tut., Vol. 16, (1), 2014.
[50] H. Zhang, C. Jiang, J. Cheng, V.C.M. Leung, Cooperative interference mitigation and handover management for heterogeneous cloud small cell networks, IEEE Wireless Commun. 22 (3) (2015) 92–99.
[51] A. Orsino, G. Araniti, A. Molinaro, A. Iera, Effective RAT selection approach for 5G dense wireless networks, 81st Conference on Vehicular Technology Conference (VTC Spring), 2015.
[52] C. Vitale, V. Mancuso, Energy efficiency in mixed access networks, in: Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2016, pp. 35–42.
[53] V. Sciancalepore, V. Mancuso, A. Banchs, et al., Enhanced content update dissemination through D2D in 5G cellular networks, IEEE Trans. Wireless Commun. 15 (11) (2016) 99 issue.
[54] M. Helou, M. Ibrahim, S. Lahoud, et al., A network-assisted approach for RAT selection in heterogeneous cellular networks, IEEE J. Select. Areas Commun. 33 (6) (2015) 1055–1067 issue.
[55] H. Zhang, W. Ma, W. Li, W. Zheng, X. Wen, C. Jiang, "Signalling cost evaluation of handover management schemes in LTE-advanced femtocell", IEEE 73rd Vehicular Technologies Conference (VTC Spring), 2011
[56] S. Wu, S. Lo, Handover scheme in LTE-based networks with hybrid access mode femtocells, J. Converg. Inf. Technol. 6 (7) (2011) 68–78 July.
[57] B. Jeong, S. Shin, I. Jang, N.W. Sung, H. Yoon, A smart handover decision algorithm using location prediction for hierarchical macro/femto-cell networks, in: IEEE Vehicular Technology Conference (VTC- Fall), 2011, pp. 1–5.
[58] S. Barmpounakis, A. Kaloxylos, P. Spapis, N. Alonistioti, Context-aware, user–driven, network-controlled RAT selection for 5G networks, Comput. Netw. J. 113 (2017) 124–147 Elsevier.
[59] J. Márquez-Barja, C.T. Calafate, J.-C. Cano, P. Manzoni, An overview of vertical handover techniques: algorithms, protocols and tools", Comput. Commun. 34 (8) (2011) 985–997 ISSN 0140-3664.
[60] Patent WO2005025260 A1, "Mobility management in mobile networks based on context information"
[61] Patent WO 2003107704 A1, "Proactive deployment of decision mechanisms for optimal handover"
[62] Patent US 20100150102 A1, "Adaptive handover mechanism for heterogeneous wireless network"
[63] Patent EP 2278840 B1, "Handover in a communication network comprising plural heterogeneous access networks"
[64] Patent EP 1670273 A1, "Handover of a mobile node between access networks of different technologies in a mobile IP telecommunications system"
[65] C. Prehofer, N. Nafisi and Q. Wei, "A framework for context-aware handover decisions," In Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2003
[66] Patent EP 1915014 A2, "Method and apparatus for handover decision by using context information in a mobile communications network"
[67] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, third ed., ISBN, 2011 9780123814791.
[68] G.W. Stewart, Matrix Algorithms Vol I, II, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.
[69] U. vonLuxburg, A tutorial on spectral clustering, Stat. Comput. 17 (4) (2007) 395–416.
[70] P.Magdalinos. D.Mavroeidis, A sequential sampling framework for spectral k-means based on efficient bootstrap accuracy estimations: application to distributed clustering, ACM Trans. Knowl. Discov. Data 7 (2) (2012) IssueAugust.
[71] The METIS 2020 Project – Laying the foundation of 5 G, 2016 [accessed April 2016]. www.metis2020.com.
[72] [accessed April 2016]. https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.1_v1.pdf.
[73] The NS3 network simulator, 2017 [accessed January 2017]. https://www.nsnam.org/.
[74] 3GPP TR 36.931 v11.0.0, "Radio Frequency (RF) requirements for LTE Pico Node B (Release 11)", September 2012.
[75] 3GPP TS 36.921 v11.0.0, "FDD Home eNode B (HeNB) Radio Frequency (RF) requirements analysis (Release 11)", September 2012.
[76] Statista, The statistics portal, 2016 [accessed April 2016]. http://www.statista.com/statistics/185868/average-mobile-wireless-call-length-in-the-united-states-since-june-1993/.
[77] Adaptive Multi-Rate Wideband, 2016 [accessed April 2016]. http://en.wikipedia.org/wiki/Adaptive_Multi-Rate_Wideband.
[78] Nokia Networks, Voice over LTE (VoLTE) optimization, 2016. White Paper available from [accessed December 2016] http://resources.alcatel-lucent.com/asset/200303.
[79] B. Theis, The great mobile network test", swisscom, 2016. available from [accessed April 2016] http://www.swisscom.ch/dam/swisscom/en/res/mobile/mobile-network/netztest-connect-en-2014.pdf.
[80] WebSiteOptimization.com, 2016 [accessed April 2016]. http://www.websiteoptimization.com/speed/tweak/average-web-page/.
[81] Virion, 2016 [accessed April 2016]. http://virion.com.au/wp-content/uploads/2010/08/VoIP-Codec-table-from-Cisco-sml.jpg.
[82] The Globe and Mail, "How much bandwidth does streaming use?", 2016. available froml [accessed April 2016] http://www.theglobeandmail.com/technology/tech-news/how-much-bandwidth-does-streaming-use/article7365916.
[83] T. Kuittinen, Skype seems miles behind WhatsApp on daily engagement, 2016. available from [accessed April 2016] http://bgr.com/2013/04/03/skype-2-billion-minutes-analysis-412280.
[84] Sysomos, Inside YouTube video statistics, 2016. available from [accessed April 2016] http://www.sysomos.com/reports/youtube.
[85] B. Genie, Mobile broadband usage guide: what can you get for your gigabyte?, 2016. available from [accessed April 2016] http://www.broadbandgenie.co.uk/mobilebroadband/help/mobile-broadband-usage-guide-what-can-you-get-for-your-gigabyte.
[86] WiFi Web Browsing Battery life, 2016. available froml [accessed April 2016] http://images.anandtech.com/graphs/graph4163/35495.png.
[87] G Web Browsing Battery life, 2016. available from [accessed April 2016] http://images.anandtech.com/graphs/graph4163/35400.png.
[88] J. Happich, Supercapacitors can take market share from lithium batteries, says IDTechEx, 2016. available from [accessed April 2016] http://www.power-eetimes.com/en/supercapacitors-can-take-market-share-from-lithium-batteries-says-idtechex.html?cmp_id=7&news_id=222905299&page=0.