Power Management in HetNets with Mobility Prediction and Harvested Energy

Abdullah Alqasir, Ahmed E. Kamal
Iowa State University (ISU), Ames, Iowa, United States,
Email: {aalqasir, kamal}@iastate.edu

***Abstract*—In this paper an optimization problem to minimize the energy drawn from the network grid by utilizing the harvested energy and dynamic sleeping of the Small Base Stations (SBSs) is presented. Due to the complexity of the optimization problem, a new UEs' movement prediction method is introduced to provide future information for the model to apply an accurate optimization problem. This method is based on a combined approach of Non-linear Autoregressive with External input (NARX) and probabilistic Latent Semantic Analysis (pLSA) to provide accurate prediction for multiple steps. Furthermore, extensive simulation results are presented to show the effectiveness of our approaches in comparison to the optimal results.**

***Index Terms*—Energy Efficiency, 5G, Energy Harvesting, Mobility Prediction, Deep Learning.**

I. Introduction

Over the past few years the cellular mobile communication technology has exponentially expanded from 2G with Small Messaging Service (SMS) to the video streaming capabilities of the 4G [1]. The main motivation behind this evolution is the rise of the data demands, where the data is predicted to grow more in the near future to reach around 400 thousands PB per month [2]. Applications like Device-to-Device (D2D)communications, Machine-to-Machine (M2M), Internet of Things (IoT), Smart cities, Health care systems and automation are emerging with needs of more robust Quality of Service QoS and scalability. The legacy systems, unfortunately are not capable of matching these new and growing demands. Therefore, 5G includes new technologies that are able to accommodate these new demands.

The SBS is a promising technology that is able to provide high rates with low energy. The SBSs are introduced to solve the problems that arise by the increasing demand for higher data rates. Therefore, they are seen as an alternative to Macro BSs (MBSs) for better coverage, quality of service (QoS) and energy efficiency. However, due to the dense deployment of SBSs, the network's energy consumption will increase as the SBSs number increases. Renewable Energy (RE) is one of the clean energy sources in next generation networks to ease the effect of climate change. Energy Harvesting is a promising solution for minimizing the energy consumption of a BS. The Harvested Energy (HE) could partially sustain the needs of the BS and for the case of SBSs could be fully sustained by HE [3].According to [4] the renewable energy has not been fully exploited in cellular networks due to economical reasons. However, with the dense deployment of the SBSs, installing HE parts could provide the network with a reliable source of energy.

Mobility prediction as part of the wireless networks design has attracted attention from academia and industry. The main concept of the mobility prediction is that giving the current and previous locations for a unique UE, what will be the next location or locations for that UE. Such predictions help optimize the wireless networks resource allocation and increase its performance. There are many research activities done in predicting the UE mobility. Works [5], [6], are based on Markov Chain, which is easy to implement but can suffer from overfitting for low data. The authors of [7], [8], are proposing mobility prediction based on Hidden Markov Model (HMM), which is more accurate than Markov chain and is able to find more complex relations between different patterns. However, HMM requires a more complex structure and higher computational capacity. The works of [9]- [13], apply the Artificial Neural Network (ANN) to predict users' mobility by learning their inherent characteristics.

Authors of [5] investigated Markov chain to predict the users movement in SBSs, specifically in Femto BSs, where their results show the prediction accuracy is affected by the regularity of the user's movement. The authors used the generated historical database for the users to discover mobility pattern and improve the prediction performance. The authors of [12] are using an enhanced Markov chain algorithm to predict user mobility by introducing an algorithm that is composed of two components: Global Prediction Algorithm (GPA) and Local Prediction Algorithm (LPA). IF GPA fails when the cell does not exist in the training database, then LPA is used. In [6], the authors employed Markov chain user prediction in proactive caching in anywhere in the network rather than just at the edge. The authors discussed a system where vehicles are connected to Roadside Units (RSU), can be BSs or Access points (APs), to allow them to connect to the internet backbone. In order to proactively cache data into the RSUs, the authors presented a mobility prediction based on Markov chain to predict the vehicle's next RSU. In [14] the authors proposed a mobility prediction based on Markov Chains to predict the users trajectory to minimize the interruption time when the handover is triggered. In [15], the authors combined the HMM with pLSA to improve the probabilistic prediction performance. The authors implemented a history-based Expectation Maximization algorithm and GPS trajectory prediction to improve the performance. Their results show successful prediction with as low as 2.2% inaccuracy. However, none of the previous works investigated a multistep prediction, which has a significant impact on the prediction quality. Therefore, we present our work which developed to produce a more accurate prediction in a multistep prediction.

The main contributions of this paper are summarized as follows:

1) We formulate an optimization problem to minimize the

energy drawn from the network grid by utilizing the harvested energy and dynamic sleeping of the SBSs. The problem updates the UEs association with SBSs in every time slot to accommodate the UEs new locations.
2) A new UEs locations prediction method is introduced which employs a combined approach of Nonlinear Autoregressive with External input (NARX) and probabilistic Latent Semantic Analysis (pLSA) to provide an accurate multistep Neural Network prediction. A new algorithm, Nonlinear AutoRegressive with external input and probabilistic latent Semantic Analysis (NARSA), is presented to show the details of the computation steps for this new joint method.
3) Lastly, we evaluate the performance of our proposed algorithm through an extensive simulation study to verify its superior computational performance compared with the optimal solution.

II. System Model and Problem Formulation

This paper considers a HetNet where several SBSs co-exist in a designated area. The SBSs are deployed randomly which provides high quality of service (QoS) for UEs.

Fig.(1) shows the architecture of the network where SBSs are equipped with energy harvesting methods (solar panels for example.) and are serving UEs under their coverage. Moreover, every SBS is connected to the Smart Grid with a two way connection.

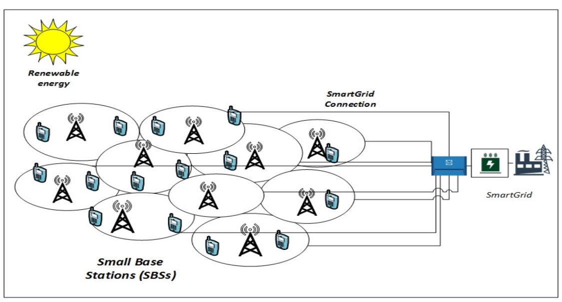

Fig. 1: The system model topology.

A. Energy Harvesting Model For SBSs

Let $f = 1, ..., F$ denote the set of the SBSs that are randomly distributed in the macro cell coverage area of $\mathbb{A}$, while $u = 1, ..., U$ and $c = 1, ..., C$ denote the set of a randomly distributed users covered by the SBSs and the set of available resource blocks in the network, respectively. Moreover, we consider a time slotted system with fixed duration τ, and $n = 1, ..., N$ denotes the index of the slot number. Furthermore, every user is assumed to be associated with only one SBS.

The SBSs harvest energy from a renewable source (e.g. wind, solar... etc), where the amount of harvested energy for every SBS f and time slot n is denoted by $hr_f[n]$ and it follows the truncated Gaussian distribution [10]. Moreover, every SBS is equipped with a battery to store its harvested energy with a maximum capacity of B_{max} with battery level at time slot n is $B[n]$. However, due to the stochastic nature of energy harvesting, every BS is connected to a non-renewable energy source to compensate for the renewable energy shortage. In other words, every SBS is set to use the energy from renewable sources first, and then request power from the grid. However, the promising technology of Smart grid which allows a two-way flow of power [11], can be used here to transfer the harvested energy between SBSs, i.e., the SBS with surplus harvested energy will transfer it to other SBSs that suffer from renewable energy deficit. Therefore, at the end of every time slot, the SBS will either transfer the surplus of its harvested energy or request energy from other SBSs to compensate its deficit. If the energy surplus of the other SBSs cannot match the energy demand of the SBS with the shortage, then the SBS will request a non-renewable energy from the smart grid directly. Hence, every SBS is equipped with two power sources: the non-renewable power from the grid and the power from the renewable sources. Therefore, the transmission power between user u and BS f using resource block c, during the time slot n is: $p^c_{fu}[n] = p^c_{fu,g}[n] + p^c_{fu,r}[n]$, where $p^c_{fu,g}[n]$ is the power drawn from the grid and $p^c_{fu,r}[n]$ is the power drawn from the renewable source (including the energy transferred from other SBSs).

Let $\lambda_f[n]$ and $\mu_f[n]$ denote the amount of the harvested energy the BS f is injecting into or receiving from the smart grid at the end of slot n, respectively. Then the amount of the harvested energy that is transferred into the smart grid equals the harvested energy that is drawn from the smart grid, where η is the transfer efficiency.

$$\mu_f[n] = \eta \lambda_f[n] \tag{1}$$

Therefore, at time slot $i = 1$ the battery will be zero, and at the end of every slot $i = 1, 2, ...N$ the battery storage will be the sum of the harvested energy subtracting the transmission power and the transferred energy $0 \leq B_f[i] \leq B_{max}$, where $B_f[i]$ is defined as:

$$B_f[i] = \sum_{n=2}^{i} hr_f[n] - \sum_{n=1}^{i} \sum_{u=1}^{U} p^c_{uf,r}[n]\tau - \sum_{n=1}^{i} \lambda_f[n] \tag{2}$$

B. User Association and Achievable Rate

Let $x^c_{uf}[n]$ be a binary indicator that is equal to 1 if user u and SBS f are associated using resource block c in n, or 0 otherwise. Also, let $z_{fu}[n]$ be a binary indicator that is equal to 1 if user u is associated with SBS f in n, or 0 otherwise. $y_f[n]$ indicates the SBS on/off status, where $y_f[n] = 0$ if the SBS is OFF during the time slot n and $y_f[n] = 1$ if the SBS is ON. However, a deactivated SBS will keep harvesting energy and injecting it into the smart grid to serve other active SBSs.

The time-varying distance between the fth SBS and the uth user can be expressed as follow:

$$d_{uf}[n] = ||l_u[n] - l_f|| \qquad \forall u \in U, \forall f \in F \tag{3}$$

where the $l_u[n]$ is the $x - y$ coordinates for the location of the user at time slot n, and the fixed location of the SBS, respectively. It follows from (3) that the channel power gain can be modeled as:

$$|h^c_{uf}[n]|^2 = \frac{\beta_0}{d^{\alpha}_{uf}[n]} = \frac{\beta_0}{||l_u[n] - l_f||^{\alpha}} \tag{4}$$

where β_0 denotes the channel gain at the reference distance of $d_0 = 1$m, and α is the path loss exponent. Moreover, the interference at a user u which is associated with SBS f from all other SBSs at a time slot n will be:

$$I_{uf}^{c}[n] = \sum_{j \neq u}^{U} \sum_{i \neq f}^{F} p_{ji}^{c}[n] |h_{ui}^{c}[n]|^2, \tag{5}$$

Then, the signal to interference and noise ratio SINR for every user is:

$$\gamma_{uf}^{c}[n] = \frac{p_{uf}^{c}[n] |h_{uf}^{c}[n]|^2}{I_{uf}^{c}[n] + \omega N_0}, \tag{6}$$

where $h_{uf}^{c}[n]$ denotes the channel gain from SBS f to user u using resource block c at time slot n, ω is the available bandwidth for every channel, and N_0 is the channel noise spectral density which is assumed to be Additive White Gaussian Noise AWGN, and ωN_0 is the noise variance σ^2. Thus, the data rate for every user using a single channel during a time slot is as follow:

$$R_{uf}^{c}[n] = \omega \log(1 + \gamma_{uf}^{c}[n]) \tag{7}$$

However, Eq.7 is non-convex, in which Taylor Expansion is used to linearize it to get the following:

$$\hat{R}_{uf}^{c}[n] = \omega \log(\sum_{u}^{U} \sum_{f}^{F} p_{uf}^{c}[n] |h_{uf}^{c}[n]|^2 + \omega N_0) - \hat{R}_{Ty} \tag{8}$$

where $\hat{R}_{Ty}$ is the first-order Taylor approximation around point ($p_{0ji}^{c}[n]$), and is as follows:

$$\begin{aligned} \hat{R}_{Ty} &\triangleq \omega log(\sum_{j \neq u}^{U} \sum_{i \neq f}^{F} p_{0ji}^{c}[n] |h_{ui}^{c}[n]|^2 + \omega N_0) \\ &+ \sum_{j \neq u}^{U} \sum_{i \neq f}^{F} \frac{|h_{ui}^{c}[n]|^2}{ln(2) p_{0ji}^{c}[n] |h_{ui}^{c}[n]|^2 + \omega N_0} (p_{ji}^{c}[n] - p_{0ji}^{c}[n]) \end{aligned} \tag{9}$$

C. Problem Formulation

In this section, an optimization problem, which minimizes the non-renewable energy consumption of the transmission power for a cooperative heterogenous network is formulated. The formulated problem evaluates users association, sleeping strategy and energy minimization within a single optimization problem. The problem can be stated as follows: given the number of users and SBSs, the problem will solve the user association, sleeping strategy and power consumption, at every time slot. The optimization problem can be mathematically state as below:

Problem $\tilde{\mathcal{P}}$:

$$\underset{\mathbb{T}}{\text{Minimize}} \sum_{f,u,n,c=1}^{F,U,N,C} p_{fu,g}^{c}[n]\tau + \sum_{f=1}^{F} E_b y_f[n]$$

subject to

$$\text{C1}: R_u^{min} \leq \sum_{f=1}^{F} \sum_{c=1}^{C} x_{fu}^{c}[n] \hat{R}_{fu}^{c}[n] \quad \forall u, \forall n$$

$$\text{C2}: \sum_{u=1}^{U} \sum_{c=1}^{C} p_{fu,r}^{c}[n]\tau \leq B_f[n-1] + \mu_f[n] \quad \forall f, \forall n,$$

$$\text{C3}: B_f[n] \leq B_{max} \quad \forall f, \forall n,$$

$$\text{C4}: \sum_{f=1}^{F} \sum_{i=1}^{n} \mu_f[i] = \sum_{f=1}^{F} \sum_{i=1}^{n} \eta \lambda_f[i] \quad \forall n$$

$$\text{C5}: \sum_{u=1}^{U} \sum_{c=1}^{C} p_{fu}^{c}[n] \leq P_f^{max} \quad \forall f, \forall n$$

$$\text{C6}: \sum_{u=1}^{U} x_{fu}^{c}[n] \leq 1 \quad \forall f, \forall c, \forall n$$

$$\text{C7}: \sum_{f=1}^{F} z_{uf}[n] = 1 \quad \forall u \forall n$$

$$\text{C8}: \frac{\sum_{c=1}^{C} x_{uf}^{c}[n]}{\#ofUs} \leq z_{uf}[n] \leq \sum_{c=1}^{C} x_{uf}^{c}[n], \quad \forall\ f, \forall u, \forall n$$

$$\text{C9}: \frac{\sum_{u=1}^{U} z_{uf}[n]}{\#ofSBSs} \leq y_f[n] \leq \sum_{u=1}^{U} z_{uf}[n], \quad \forall f, \forall n$$

where $\mathbb{T} = \{p_{uf}^{c}[n], \lambda_f[n], \mu_f[n], y_f[n], z_{uf}[n], x_{uf}^{c}[n]\}$. Constraint (C1) represents the QoS for every user. The constraints from (C2) to (C5) are dealing with energy transfer and cooperation between SBSs. Constraint (C2) represents the energy consumption causality where the BS cannot use energy more than what is available. Constraint (C3) limits the battery capacity. Constraint (C4) is for energy conservation, where the total injected energy into the smart grid equals the total received energy by all BSs. Constraint (C5) limits the maximum allowed transmission power for every BS. Constraints from (C6) to (C9) are dealing with the UEs association and SBSs sleeping strategy.

Problem $\tilde{\mathcal{P}}$ is an Mixed Integer NonLinear Problem (MINLP) which is too complex due to the coupling of the binary and continuous variables. Moreover, due to the dynamic nature of UEs movement, where each UE is expected to move in every time slot which causes the system to update its association and sleeping strategy accordingly, problem $\tilde{\mathcal{P}}$ is required to be solved every time slot to match the changes.

One solution for this problem is to predict the future location of the UEs and solve the problem according to the predicted location of each UE. Therefore, a user mobility prediction approach is introduced , where the UEs' movements are predicted for the next N slots. This will help simplify the problem where it will be solved for the next N predicted time slots instead of every slot.

III. Prediction Model for User Mobility

In this section we investigate two approaches to predict the UEs' future mobility pattern and location. First, we solely apply Nonlinear Autoregressive (NAR) method to predict each UE individually. Second, a joint approach to predict the UEs' future mobility pattern and location by exploiting ANN Nonlinear Autoregressive with External input (NARX) and probabilistic Latent Semantic Analysis (pLSA) to provide an accurate multistep Neural Network prediction. This joint approach leverages ANN NARX outstanding results in time series prediction tasks and the (pLSA) ability in detecting hidden patterns between different UEs.

A. Nonlinear Autoregressive (NAR) Time Series Prediction

The Nonlinear Autoregressive (NAR) are types of Neural Networks that are used to forecast samples framed in a onedimensional time series. NAR networks is used to predict the value of certain time series data by using the past data. The following equation shows how NAR networks work:

$$y(n) = f(y(n-1), y(n-2), ..., y(n-d)) \tag{10}$$

where the function f is unknown and the neural network is used to approximate. This equation describes how NAR works in predicting the future values of y. In every step, NAR uses the past d values of y to predict the future data point in series y.

The advantage of NAR is its simple structure and requires less computation time. However, for multistep prediction where the predicted output $\hat{y}(n)$ is used to predict future results, the error will adds up in every step which is accumulated to produce inaccurate results. Therefore, NARX is better in predicting future works since it relies not only on the data but on other external set of correlated data.

B. Nonlinear Autoregressive with External input (NARX) Time Series Prediction

In many applications, there are an important correlation between the predicted time series and some other external data. Some stock exchange prices are correlated with certain times of the years, e.g., Apple stock price after the holidays. The following equation shows how NARX networks work:

$$y(n) = h(y(n-1), y(n-2), ..., y(n-d), k(n-1), ..., k(n-d)) \tag{11}$$

Similar to NAR networks, NARX predicts the future of y according to the past d values, where for every one future prediction NARX will employ the past data to make its prediction. However, NARX includes the external data set $k(n)$ to approximate the function h.

NAR is simpler to perform and requires less computation power. However, NARX can have better performance, specially when there is strong correlation between the predicted data and the external data. Therefore, in the following section a statistical method is employed to detect the correlation between different data sets to use them in performing the prediction.

C. probabilistic Latent Semantic Analysis (pLSA)

pLSA was originally introduced to derive a representation of the observed variables in terms of certain hidden variables [16]. pLSA is a statistical technique used for the analysis of co-occurrence data which is based on a mixture decomposition derived from a latent class model. Mobile UEs usually follow daily movement patterns, e.g., some UEs use public transportation daily, others drive to business districts. Thus, pLSA is used in this work to reveal the hidden patterns that each UE is using and employ this pattern as an external data, i.e., $k(n)$ in the previous section, to provide more accurate predictions.

Let us associate an unobserved class variable $m_k \in m_1, m_2, ..., m_K$ with each observation, where the observation is the user's movement from one location to another. Moreover, let us denote the user's location at time slot n to be equal to $l[n] \in \mathbb{A}$, where, $\mathbb{A}$ is a predefined area. The user's movement from location a to location b is denoted by $l^a[n] \rightarrow l^b[n]$. Here, we are considering the direction of the user's movement without considering the traveled distance. Thus, we denote $\mathcal{L}[n]$ as the direction of the movement during time n. Hence, the probability of the movement toward direction $\mathcal{L}[n]$ is:

$$P(\mathcal{L}[n]) = \frac{\text{Number of movements toward direction } \mathcal{L}[n]}{\text{Total number of movements}} \tag{12}$$

Moreover, every observed data item is a pair of data $(u, \mathcal{L}[n])$. Therefore, the joint probability of the observed data will be:

$$P(u, \mathcal{L}[n]) = \sum_{k=1}^{K} P(u)P(m_k|u)P(\mathcal{L}[n]|m_k) \tag{13}$$

where parameter $m_k \in [m_1, m_2, ..., m_K]$ denotes the mobility class, and the probability that a user u is following mobility class m_k at any given time, is defined as $P(m_k|u)$. Moreover, the probability of the direction of movement given the mobility class m_k is $P(\mathcal{L}[n]|m_k)$, while $P(u) = \frac{m(u,\mathcal{L}[n])}{\sum_{u=1}^{U} m(u,\mathcal{L}[n])}$ denotes the probability that u made a movement, where $m(u, \mathcal{L}[n])$ is the number of times UE u made a movement.

However, the two probabilities $P(\mathcal{L}[n]|m_k)$ and $P(m_k|u)$ cannot be calculated analytically, since the classes m_k are unknown. The Expectation Maximization (EM) algorithm is a well-known algorithm that is used to compute Maximum Likelihood Estimates (MLE) [17]. The EM algorithm consist of two consecutive steps: an expectation step, followed by a maximization step. The first step: the expectation (E) step where posterior probabilities are calculated for the latent variables, based on the current estimates of the parameters. The second step the Maximization (M) step where parameters are updated to maximize the expected complete data log-likelihood, which depends on the posterior probabilities computed in the E step.

In our model the classes $m_k \in M = \{m_1, m_2, ..., m_K\}$ are unobserved and unknown, which can only be estimated using the EM algorithm. Applying the EM algorithm to the pLSA model as in [16] introduces the two steps as follows: The expectation (E) step the Bayes' estimator is calculated for the latent variables, based on the current values of $P(m_k|u)$ and $P(\mathcal{L}[n]|m_k)$. The Maximization (M) step is used to update the parameters to maximize the expected complete data log-likelihood, which depends on the $P(m_k|u)$ and $P(\mathcal{L}[n]|m_k)$ equations that are computed in the E step.

In the Expectation step, we use Bayes' estimator to calculate the posterior probabilities based on the current estimates of the parameters. The Bayes' estimator is as follows:

$$P(m_k|u, \mathcal{L}[n]) = \frac{P(\mathcal{L}[n]|m_k, u)P(m_k|u)}{\sum_{m \in \mathbb{M}} P(\mathcal{L}[n]|m_k, u)P(m_k|u)} \tag{14}$$

In the Maximization step, the expectation of the complete data log-likelihood $\mathbb{E}[\mathbb{L}]$ is maximized as follow:

$$\begin{aligned}\mathbb{E}[\mathbb{L}] = \sum_{u=1}^{U} \sum_{\mathcal{L}[n]} m(u, \mathcal{L}[n]) * \\ \sum_{k=1}^{K} P(m_k|u, \mathcal{L}[n]) log\big[P(\mathcal{L}[n]|m_k, u)P(m_k|u)\big]\end{aligned} \tag{15}$$

where $m(u, \mathcal{L}[n])$ indicates the number of times the user moved according to direction $\mathcal{L}[n]$ during time slot n. Then, the following two re-estimation equations are used in the M-step:

$$P(\mathcal{L}[n]|m_k) = \frac{\sum_u m(u, \mathcal{L}[n]) P(m_k|u, \mathcal{L}[n])}{\sum_{n=1}^{N} \sum_u m(u, \mathcal{L}[n]) P(m_k|u, \mathcal{L}[n])} \quad (16)$$

$$P(m_k|u) = \frac{\sum_u \sum_{n=1}^{N} m(u, \mathcal{L}[n]) P(m_k|u, \mathcal{L}[n])}{\sum_{n=1}^{N} \sum_u m(u, \mathcal{L}[n])} \quad (17)$$

After evaluating Eqs. (16) and (17) each UE will be assigned to the class with the highest probability, which it will share with other UEs where they share a common pattern. Let us denote α_k^u as the UE u that belongs to class k and Ω_k as the group that contains all UEs that belong to class k. Lets denote α_k^* as the UE that has the highest probability in class k.

In the prediction stage using NARX network, every group Ω_k will choose the UE that has the highest probability in $P(m_k|u, \mathcal{L}[n])$ as the external data $k(n)$ for predicting the movement of every UE in that Group. This ensures the external data that is used in NARX has a strong correlation with all of its associates.

Algorithm 1 UEs' Movement Prediction using NARSA.

Input: $u; \mathcal{L}[n]; K; \epsilon$
Initialize $P^0(m_k|u)$;$P^0(\mathcal{L}[n])|m_k)$; $\Delta \leftarrow \infty$;$i \leftarrow 1$
while $\Delta \geq \epsilon$ **do**
  Compute $P^{[i]}(m_k|u, \mathcal{L}[n])$ in (14) using $P^{[i-1]}(m_k|u)$ and $P^{[i-1]}(\mathcal{L}[n])|m_k)$;
  Update $P^{[i]}(m_k|u)$, $P^{[i]}(\mathcal{L}[n])|m_k)$ using (16), (17);
  Compute the expectation $\mathbb{E}[\mathbb{L}]^{[i]}$ in (15);
  Δ = $|\mathbb{E}[\mathbb{L}]^{[i]} - \mathbb{E}[\mathbb{L}]^{[i-1]}|$;
  $i \leftarrow i + 1$;
end while
Output: $\mathbf{P}(m_k|u)$, $\mathbf{P}(\mathcal{L}[n])|m_k)$,$\alpha_k$,$\alpha_k^*$,$\Omega_k$.
for k=1:K **do**
  $k(n) \leftarrow \alpha_k^*$
  for $\forall$ UE $\in \Omega_k$ **do**
    $y(n) \leftarrow \alpha_k^u$
    Train NARX Network by estimating Eq.11
  end for
end for
Output: $\hat{y}(n)$

Algorithm 1: NARSA is constructed of two parts: The first is the EM algorithm, and the second is the NARX network training. First we initialize the probabilities of $P(\mathcal{L}[n])|m_k)$ and $P(m_k|u[n])$ with random initial values, then the algorithm will alternate between the Expectation and maximization parts i.e., between Eq.(14) and Eqs.(16) and (17). In every step $\mathbb{E}[\mathbb{L}]^{[i]}$ is calculated using eq.(15), and compared to $\mathbb{E}[\mathbb{L}]^{[i-1]}$. If the difference is less than Δ the algorithm will stop and $\mathbf{P}(m_k|u)$, $\mathbf{P}(\mathcal{L}[n])|m_k)$ will be the local optimal values and α_k^* is computed. Otherwise the algorithm will repeat the previous two steps again. At the second part, NARSA will employ the results from EM to feed NARX network the external data and train it to predict $\hat{y}(n)$. However, since the EM is nonconvex, the algorithm is repeated with different initials multiple times to find the maximum value for $\mathbb{E}[\mathbb{L}]$.

Table I: Simulation Parameters

Parameter	Value	Parameter	Value
P_{max}	4 watts	R_{min}	1 Mbps
N_0	-174 dbm/Hz	ω	5 MHz
B_{max}	600 Joules	τ	10s
η	0.9	E_b	20 Joules

IV. Simulation Results

In this section we evaluate the performance of algorithm 1, on the Mobile Data Challenge (MDC) data set [18] and [19]. DMC data contains GPS traces for both pedestrians and vehicular from Lake Geneva region in Switzerland. The data is gathered from participants using GPS where their locations are recorded every 10 seconds for over one year of time. The data is processed to improve their quality, i.e., some outliers are removed. Figures 2 and 3 evaluate the performance of Algorithm 1 to predict UEs' movements and compare it with NAR method. The system parameters are listed in Table I unless stated otherwise.

Fig.2 shows the comparison between NAR and NARX. In this figure we use an ANN with three layers, the first is the input layer, the second is the hidden layer and the third is the output layer. The hidden layer consists of 10 neurons with Tanh function. In this part we applied a delay of 4 steps, i.e., $d = 4$. Moreover, 10 UEs are used from DMC data set with $m_k = 3$ classes. One UE data from every class is used as an external data in NARX, while NAR system does not require external data. Moreover, 2000 data pairs are used to train the network, with 70% for training, 15% for validation and 15% for testing purposes. The figure shows the accumulated error percentage of both NAR and NARX systems. As shown in the figure NAR predicts the next two steps with very accurate prediction, however, after the fifth prediction NAR accumulated error, which is defined as how far the prediction is compared to the actual location, increases rapidly. On the other hand, NARX kept a relatively accurate prediction until the tenth step with error percentage of around 10%. This is understandable since the external data has a strong correlation with the input data that is being predicted.

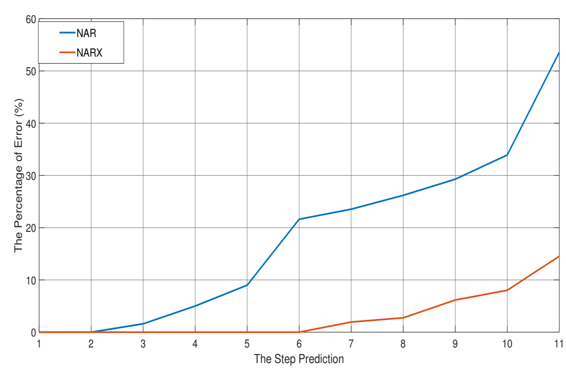

Fig. 2: Comparison between NARX and NAR for 11 Steps.

With the same setting as in Fig. 2, Fig. 3 show the probability of success as the number of prediction increases. In this figure we also compare NAR with NARX, and as the figure depicts both approaches start with high probability of success but as the number of prediction steps increases the performance of NAR decreases rapidly, while the performance of NARX decreases slowly. This is understandable since the highly correlated external input that is used in NARX helped sustaining the prediction for more steps.

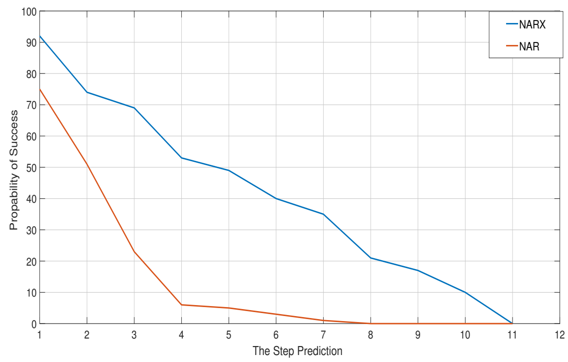

Fig. 3: The probability of success as the prediction steps increases.

Fig 4 shows the change of power consumption due to the prediction inaccuracy. In this figure we solved problem $(\widetilde{\mathcal{P}})$ according to the actual locations and predicted locations of both NAR and NARX. We used 8 SBSs that are placed to cover the area of the future movement of 1 UE. First the problem was solved according to the actual locations in order to decide the user association in every time slot n. Then, using this association we changed the UEs locations according to the prediction to calculate the consumed power. The figure shows how the inaccurate prediction affected the energy consumption in both NAR and NARX. In NAR, after the fifth step the assigned SBS to the UE cannot maintain the communication since it used the maximum allowed power, which is represented by the straight red line. On the other hand, NARX has a more accurate prediction of the system that kept the UE associated with SBS until the tenth step.

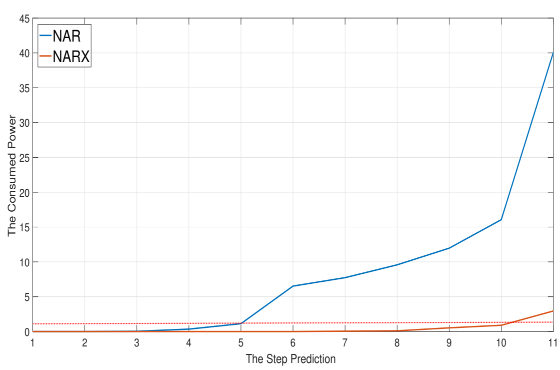

Fig. 4: Consumed Power Due to Prediction Error.

V. Conclusions

In this paper, an optimization problem is formulated to minimize the energy drawn from the network grid, with updating the UEs association with SBSs in every time slot instead of keeping the association stationary. A new UEs' prediction method is introduced that is based on a combined approach of Non-linear Autoregressive with External input(NARX) and probabilistic Latent semantic Analysis (pLSA) to provide accurate prediction for longer times. Moreover, a new algorithm NARSA is presented to show the general steps that are employed to predict the UEs' next locations. The results show that the employment of UEs prediction and ANN in solving the optimization problem gave efficient computational performance with a near optimal solution.

References

[1] S. Sarraf, "5g emerging technology and affected industries: Quick survey," American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS), vol. 55, no. 1, pp. 75–82, 2019.

[2] C. V. N. I. Cisco, "Global mobile data traffic forecast update, 2017–2022," white paper, 2018.

[3] G. Piro, et al., "Hetnets powered by renewable energy sources: Sustainable next-generation cellular networks," IEEE Internet Computing, vol. 17, no. 1, pp. 32–39, 2012.

[4] G. Piro, et al., "Hetnets powered by renewable energy sources: Sustainable next-generation cellular networks," IEEE Internet Computing, vol. 17, no. 1, pp. 32–39, 2013.

[5] N. A. Amirrudin, et al., "User's mobility history-based mobility prediction in lte femtocells network," in 2013 IEEE International RF and Microwave Conference (RFM), pp. 105–110, Dec 2013.

[6] N. Abani, T. Braun, and M. Gerla, "Proactive caching with mobility prediction under uncertainty in information-centric networks," in Proceedings of the 4th ACM Conference on Information-Centric Networking, pp. 88–97, ACM, 2017.

[7] K. Yap and Y. Chong, "Optimized access point selection with mobility prediction using hidden markov model for wireless network," in 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), pp. 38–42, July 2017.

[8] Y. Qiao, et al., "A hybrid markov-based model for human mobility prediction," Neurocomputing, vol. 278, pp. 99 – 109, 2018.

[9] D. S. Wickramasuriya, et al., "Base station prediction and proactive mobility management in virtual cells using recurrent neural networks," IEEE 18th Wireless and Microwave Technology Conference (WAMICON), pp. 1–6, April 2017.

[10] A. Alsharoa, H. Ghazzai, A. E. Kamal and A. Kadri, "Optimization of a Power Splitting Protocol for Two-Way Multiple Energy Harvesting Relay System," in IEEE Transactions on Green Communications and Networking, vol. 1, no. 4, pp. 444-457, Dec. 2017.

[11] X. Fang, S. Misra, G. Xue and D. Yang, "Smart Grid — The New and Improved Power Grid: A Survey," in IEEE Communications Surveys & Tutorials, vol. 14, no. 4, pp. 944-980, Fourth Quarter 2012.

[12] A. Hadachi, et al., "Cell phone subscribers mobility prediction using enhanced markov chain algorithm," Intelligent Vehicles Symposium Proceedings, pp. 1049–1054, June 2014.

[13] Y. Tang, et al., "Delay-minimization routing for heterogeneous vanets with machine learning based mobility prediction," IEEE Transactions on Vehicular Technology, vol. 68, pp. 3967–3979, April 2019.

[14] A. Mohamed, et al., "Mobility prediction for handover management in cellular networks with control/data separation," in Communications (ICC), pp. 3939–3944, IEEE, 2015.

[15] D. Stynes, K. N. Brown, and C. J. Sreenan, "A probabilistic approach to user mobility prediction for wireless services," in Wireless Communications and Mobile Computing Conference, pp. 120–125, IEEE, 2016.

[16] T. Hofmann, "Probabilistic latent semantic indexing," in Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '99, (New York, NY, USA), pp. 50–57, 1999.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," Journal of the Royal Statistical Society, vol. 39, no. 1, pp. 1–38, 1977.

[18] J. Laurila, et al., "The mobile data challenge: Big data for mobile computing research. nokia research center," 01 2012.

[19] N. Kiukkonen, B. J., O. Dousse, D. Gatica-Perez, and J. K. Laurila, "Towards rich mobile phone datasets: Lausanne data collection campaign," in Proc. ACM Int. Conf. on Pervasive Services, Berlin, July 2010.