A Novel Approach for Cluster Self-Optimization Using Big Data Analytics

Abbas Mirzaei*
Department of Computer Engineering, Islamic Azad University, Ardabil Branch, Ardabil, Iran
a.mirzaei@iauardabil.ac.ir
Amir Rahimi
Department of Computer Engineering, Islamic Azad University, Ardabil Branch, Ardabil, Iran
dramr@ihemardabili.ac.ir

Received: 05/Feb/2019 Revised: 28/Aug/2019 Accepted: 28/Sep/2019

Abstract

One of the current challenges in providing high bitrate services in next generation mobile networks is limitation of available resources. The goal of proposing a self-optimization model is to maximize the network efficiency and increase the quality of services provided to femto-cell users, considering the limited resources in radio access networks. The basis for our proposed scheme is to introduce a self-optimization model based on neighbouring relations. Using this model, we can create the possibility of controlling resources and neighbouring parameters without the need of human manipulation and only based on the network's intelligence. To increase the model efficiency, we applied the big data technique for analyzing data and increasing the accuracy of the decision-making process in a way that on the uplink, the sent data by users is to be analyzed in self-optimization engine. The experimental results show that despite the tremendous volume of the analyzed data – which is hundreds of times bigger than usual methods – it is possible to improve the KPIs, such as throughput, up to 30 percent by optimal resource allocation and reducing the signaling load. Also, the presence of feature extraction and parameter selection modules will reduce the response time of the self-optimization model up to 25 percent when the number of parameters is too high Moreover, numerical results indicate the superiority of using support vector machine (SVM) learning algorithm. It improves the accuracy level of decision making based on the rule-based expert system. Finally, uplink quality improvement and 15-percent increment of the coverage area under satisfied SINR conditions can be considered as outcome of the proposed scheme.

Keywords: Self-Optimization Networking; Big Data; Quality of service (QoS); Resource Allocation; Load Balancing.

1. Introduction

Over the last four years, annual mobile data traffic has increased more than 130%. It is predicted to continue to increase from 2016 to 2020 at an accumulated annual rate of 78%. Despite the technical and economic advantages of using femto-cell layer in new generation of mobile network, establishing the femto-cell layer for the operators entails some challenges. In order to reach the expected spectrum efficiency, the macro and femto layers must operate at a single bandwidth, which will make interference more erratic and its control more difficult. The responsibility of the resource management center (RMC) will become more complex than for conventional cellular networks. In this regard, SON is introduced as the only viable solution for controlling and managing this type of huge data networks [1].

The control and management of various nodes in mobile networks require self-organization and smart organization.

2. Related Works

The most recently-proposed structures for current mobile networks are combinations of the functions of the base station [4]. These products require a data connection such as DSL through which to connect to the main network of the mobile operator. Its capabilities should include a maximum transfer strength of about 20 dBm, coverage of high-speed packet access and the ability to establish a certain number of simultaneous voice calls or data sessions. Basic functions of the PnP, such as self-configuration and activation, are carried out in a very limited manner.

Researchers have proposed numerous methods based on automatic neighbour relations (ANR) for optimization in large scale networks. A number of schemes have been developed based on the selection of automatic neighbour relations [5], [6], [7].

Authors [8] proposed a solution for increasing the quality of the service which is to reduce the number of periodical errors during connection. Reducing the number of drop rates and increasing the HOSR increases the QoS in the cellular networks. The authors proposed a framework for determining the neighbour cell list (NCL) to increase the call maintenance indicator and the HOSR. They showed that improving the efficiency of the network relates directly to the NCL. One problem with this scheme is the imbalance in the level of cell coverage with the changes in the NCLs. This has no significant influence under ordinary conditions, but can lead to failure of service provision.

* Corresponding Author

In [9] Authors found that an automatic NCL scheme is more reliable than similar schemes. It is implemented by adding a management register called a network management system to the network core to allow scalability of the network. When a new entity is added to the network, it connects without problem and can be introduced to other parts of the network through a logical procedure during service provision.

In [10] a scheme for third-generation mobile networks has been considered where the requirements of load balancing are considered in the design phase. The authors considered the transmission power of the downlink as well as measures which influence the effective coverage area of the cell. In this scheme, prediction error due to the fallibility of GIS and the probability of incorrect mapping of the data can increase the rate of unsuccessful handovers. The high fallibility of this scheme means it cannot be considered a highly efficiency pattern.

Authors in [11] proposed a scheme where UTRA FDD-based networks provide good ability for configuring and adjusting inter-frequency handovers between cells based on assessment of the NCLs. While this scheme is advantageous from some aspects, the failure to consider inter-frequency handovers in next-generation mobile networks decreases its value to a great degree. The scheme proposed in [12] introduces big data technology to optimize fifth-generation mobile networks. Despite the suitability of this technique for next-generation mobile networks, it is only loosely introduced and no output is provided to compare with similar schemes.

Some authors [13] proposed a self-optimization scheme for optimizing neighbour relation indicators by covering the gaps present in the coverage area of the network. Despite the positive and efficient aspects of this scheme, it only considers optimization for determining ANR. As seen from the algorithms in this scheme, it can only be used under static conditions to increase the level of successful neighbour connections between cells. Assessments show that the previous patterns have some disadvantages. The next section describes our proposed model, which addresses the shortages seen in previous schemes.

The proposed approach is a self-optimization model based on capacity/coverage index management by controlling the network parameters and considering events, timing factors, limited network resources and population distribution of the coverage area. The proposed self-optimization engine comprises four main modules. Big data technology is applied to make the proposed model smart and increase its decision-making capability. The challenge of managing big data in mobile networks has been significantly addressed. Extracting useful and structured data requires the use of a great chunk of information in a very short timeframe and the ability to extract meaningful information from insignificant data. Reaching a logical relationship between telecommunication and statistical parameters and events can improve decision-making in the mobile network about the use of power and frequency resources in the handover among components.

3. The Move Towards Distributed Intelligence: Big Data based SON

The basis of the proposed approach is to provide a self-optimization model which allows for control of resources and measures related to neighbourhood relations in the mobile network without human interference and by reliance solely on network intelligence. To increase efficiency, big data techniques will be used for the data analysis and decision-making of the network regarding resource allocation. The need to optimal resource supply in areas covered by the network and ensuring the user QoS require control enormous volume of data and manage the configuration characteristics of new generation wireless networks [14]. This will result in an exponential increase in the functional complexity of the design and optimization for this type of network. Self-optimization techniques are the only viable solution for controlling and managing this type of networks that allows control of resources and key performance indicators (KPI) of the mobile network without human intervention based solely on the intelligence of the network [15]. The main objective behind such a self-optimization model is to maximize network efficiency and increase the quality of service (QoS) provided to macro-cell and femto-cell users with the limited network resources. Although some methods such as Learning, Fuzzy Logic, and Convex Optimization approaches have been used to make self-optimization models intelligent but in next-generation networks, such non-intelligent methods don’t provide acceptable functionality because of their computational complexity restrictions. Also, their computational complexity depends to scale of network and it will increase dramatically with the increasing volume of data exchanged. Data in these networks require high volume, high speed, and variety, which are the main features of a powerful data management technology called big data so, this study proposes an approach that uses a self-optimization framework based on this technology [16].

In order to satisfy the QoS requirements in multimedia services, resource allocation algorithms must be modified and adjusted to clarify the distinction among services and network conditions [17]. Previous studies have shown that demand side management increases the efficiency and effectiveness of the network to supply resource demand considering user utility and cost [18]. A proper efficient resource allocation based on peak reduction has been presented by [19] in which reducing the network operational cost was considered, as well as resource supplier's cost. The optimality of most current distributed optimization schemes are based on achieve a single objective for example utility of user side does not guarantee satisfaction of hard QoS constraints, although; they rely on optimization schemes to apply the proper scheme to minimize cost functions [20], [21].

4. BD-SON Configuration

The way to utilize the self-optimization technology in next-generation mobile networks is through a specific change in the architecture of these networks. The first section added to the structure of the networks is to the data gathering unit, which must be able to receive data from parts and sections and classify the data. The data gathered must be filtered to extract the correlation and meaning and that with the more important meaning will be entered into the self-optimization engine [22]. As you can see in the architecture of the proposed self-optimization networking model (Figure 1), we need to apply at least two additional section through the architecture of current generation mobile networks. The first register by the name of Big Data Gateway is responsible for collecting data from Femto-cell level and send it to the core for further evaluation. Also, another register is Big Data SON Engine which will be applied to making decision based on network condition and the level of resources and customer behaviors which it is in communication with some other parts of the core registers such as MME and S-GW. You can take a look at the introduced architecture as the Figure (1):

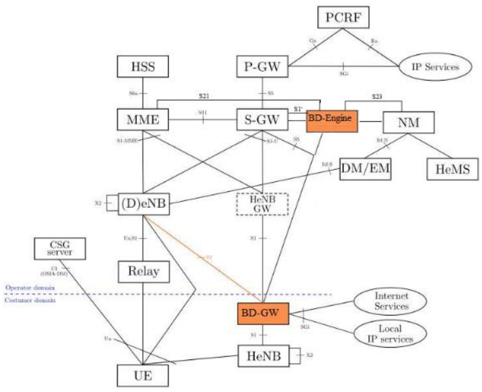

Figure 1. The modified structure of NGMN using BD-SON

The proposed scheme provides a framework for overcoming challenges to 4G and 5G networks using self-organizing networks (SONs) which is consistent with the characteristics of next-generation multi-carrier mobile networks. A general framework for reinforcing the SON using big data is proposed to improve the decision-making resource of the intelligent self-optimization system and the general efficiency of the network. Different data analysis techniques are used and pre-processing before decision-making are the logical methods for decreasing the state space of decision-making and the response time.

We have introduced the configuration of the proposed Self-Optimization Networking in previous part. As mentioned, automatic controller engine and the resource controller are the most important part of all self-organized networks. So, in order to more accurate investigation of the functionality of this model, we describe all parts of the proposed model in detail. In this regard, you can find the various characteristics and their interfaces in the Figure (2).

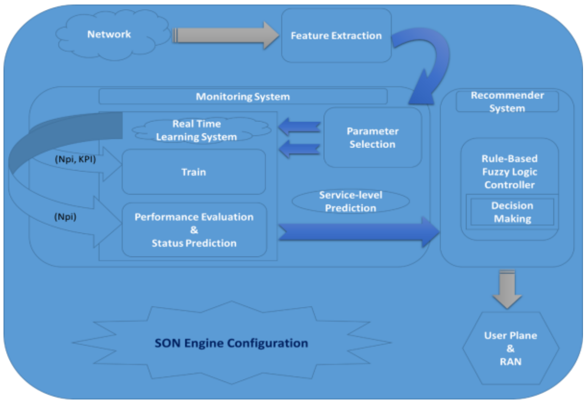

Figure 2. The component of SON Engine

After analyzing the input data and making decision, we have two external command follow which the first commands is relevant to Resource Management Center (RMC) for adjustment of resource level of each network entities and another command sequences which applied to modification of all decision-making thresholds and various applied timing factors. Also, in order to describe the SON configuration better, we should consider the relation of engine with other parts of the radio access networks.

Figure 3 demonstrates the functioning of the self-optimization model. The flowchart records the data from the access network layer, events and user-level data in a register. After the pre-processing phase, which involves the classification of the data, filtration, rating and checking the level of correlation between data, the data enters the efficacy assessment unit. In this unit, the status of the network for a certain time period is predicted and, if the efficiency levels are satisfactory, after a time delay, it returns to monitoring the status of the network. If the efficiency levels are not satisfactory, the self-optimization engine will be activated. After analysis of the proven relations between the efficiency indicators and the network parameters, network decision-making is initiated. The results are sent to the RMC. The control information transferred between the self-optimization engine and the RMC includes details about the allocation of resources to various components of the network for a specific time period in the future.

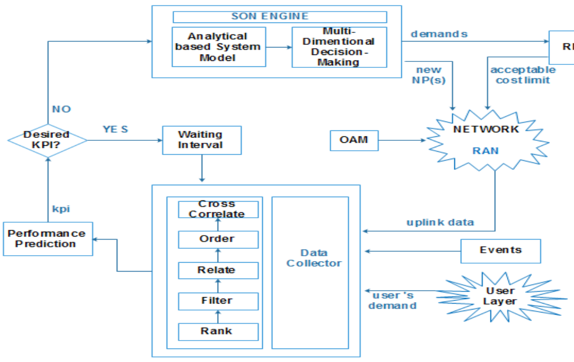

Figure 3. The proposed BD-SON architecture

In the RMC, the response to the self-optimization engine request is provided such that resource allocation accrues the lowest possible cost. The working algorithm attempts to optimally allocate resources based on requests from the SON engine using a pre-defined function. The control messages related to resource allocation and commands for readjustment of the threshold levels in neighbour relations are sent to the MME unit of the network core. The resource allocation commands are executed in accordance with the handover operators in the network.

5. Optimal Resource Allocation in the Distributed Self-Optimization Model

One of the factors influencing the level of service quality is the ideal allocation of the resources based on the network's load distribution approach. In order to minimize the load variance, the proposed self-optimization model utilizes an intelligent resource allocation pattern. Moreover, while allocating resources to users, cost limitations and the upper limitation of the total power of the network must also be considered.

In this part of the assessment for the proposed self-optimization model, the level of efficiency for femto-cell networks based on the proposed method and other resource allocation patterns will be evaluated. Accordingly, the quality of communication channels regarding the SINR indicator, the coverage area, the throughput, and the uplink/downlink quality will be discussed.

In order to use this power allocation framework in the self-optimization model, the resource distribution approach is presented as a three-part online algorithm that converges to optimal values after a number of iterations. The various steps of the algorithm are as follows: The constant coefficients related to the optimization relations based on the current conditions of the user and the received service are determined through the analysis of uplink data. Using this approach in the resource allocation process in the self-optimization model, the quality requirements of the users will be met by considering the costs and with a low computational load.

Ideal resource allocation based on the network load distribution is a main QoS affecting factor [23]. In order to minimize load variance, the proposed self-optimizing model utilizes an intelligent resource allocation pattern. Cost constraints and the upper bound of network power limitation must be considered when distributing resources among entities. In order to allocate power to users over time intervals $t = [1, 2, \dots, T]$, information related to flexibility $\omega_i(t)$ and cost component $c(t)$ must be determined by the big data processing engine through analysis of the received data. $P_i(t)$ denotes the power consumption by user i at moment t and P denotes the restriction on network total power. The problem of real time resource assignment can be formulated as follows:

$$max: \sum_{i=1}^{T}\sum_{i\in N}\left[U(P_i(t),\omega_i(t)) - f\left(\sum_{i\in N} P_i(t)\right) P_i(t)\right] - \frac{\alpha T}{2} Var\left(\sum_{i\in N}\vec{P}_i\right) \quad (1)$$

where:

$$P_i(t) \geq P_{i,min}(t), \forall i \in \mathbb{N}, t \in \{1,2,\dots,\mathrm{T}\} \quad (2)$$

$$C\left(\sum_{i\in \mathbb{N}} P_i(t)\right) \leq c(t), \forall t \in \{1,2,\dots,T\} \quad (3)$$

where $\mathrm{Var}(.)$ denotes the variance of the network total power, $\mathrm{f}(.)$ denotes the cost function and $\mathrm{U}(.)$ denotes the utilization function. To provide consistency between the network load variance and satisfaction of user QoS constraints, parameter α is defined as a positive value. $P_{i,min}(t)$ is the user's i^{th} minimum power request at t. In this framework, N denotes the node set of the network.

Formulation (1) expresses the goal function of the model. Actually, this formulation is included a utilization function denoted by $U(P_i(t), \omega_i(t))$ a cost function which is shown as $f(\sum_{i\in N} P_i(t))\, P_i(t)$ in addition to load balancing part which has been formulated based on variance function $(\frac{\alpha T}{2} Var(\sum_{i\in N} \vec{P}_i))$.

The proposed approach is expressed in the form of a three-step optimization model which has a unique solution. This real-time algorithm converges to optimal values after a number of iterations as follows:

Step 1: Begin $\forall i \in \mathbb{N}$, initialize $\widehat{P}_i(0) \in \mathbb{P}$.

Step 2: During consecutive iterations: for time slot t, the below convex optimization problem has been solved by SON engine (mentioned Probe-On):

$$max: \sum_{i\in N} U(p_i(t),\omega_i(t)) - f\left(\sum_{i\in N} p_i(t)\right)\sum_{i\in N} p_i(t) - \frac{\alpha}{2}\sum_{i\in N}\left(p_i(t) - \hat{p}_i(t-1)\right)^2 \quad (4)$$

where:

$$p_i(t) \geq p_{i,min}(t), \forall i \in \mathbb{N} \quad (5)$$

$$C\left(\sum_{i\in\mathbb{N}} p_i(t)\right) \le c(t), \forall t \in \{1,2,\dots,T\} \qquad (6)$$

Which $\widehat{P}_i(0) \in \mathbb{P}$ shows the set of primary power vector relevant to all network's entities. Note that the sum of the dedicated powers should not be more than the limitation of the upper bound of the power. The solution of Probe-On is shown by $\vec{p}^*(t)$, where in these formulation $p_i^*(t)$ indicates the optimal assigned power to user i.

Step 3: In each time slot t for all $\forall i \in \mathbb{N}$, Update $\hat{p}_i(t)$ using following equation:

$$\hat{p}_i(t) = \hat{p}_i(t-1) + \frac{\alpha}{t+\alpha} \cdot \left(p_i^*(t) - \hat{p}_i(t-1)\right) \quad (7)$$

The mentioned coefficients i.e., ω and α are determined based on current user status, service type and network condition through uplink data analysis. Using this method for resource allocation, the self-optimization model satisfies QoS constraints with the minimum cost and low computational load.

6. Distributed Load-Balanced Resource Allocation

Determining the optimal values in the proposed model is not purely based on local node variables. This method solves optimization Lagrange equations iteratively; each node calculates the updated Lagrange coefficients and uses them for the next iteration of the power allocation algorithm. In the system model, it is assumed that the system has N users, and the K^{th} user has a data rate equal to R_k bits per symbol. $C_k^{(c)}$ is the number of bits allocated to the c^{th} class-carrier for the K^{th} user. In the transmission channels, different class-carriers will experience different channel gains, denoted by $\alpha_k^{(c)}$, the magnitude of the c^{th} class-carrier seen by the K^{th} user. It is assumed that N_0 is white noise and is equal for all class-carriers and the same for all users. The goal of this approach is the best assignment of $C_k^{(c)}$ which besides of satisfying total power constraint, the capacity of links has been maximized. Note that this problem can be formulated either to minimize the transmission powers besides satisfying the given QoS requirements or to improve the user QoS parameters for a fixed overall transmission power. The formulation for the second approach can be achieved by changing the class-carrier power levels proportionally using the same set of $C_k^{(c)}$. The enhancement in quality can be demonstrated by the increase in total user transmission rate (R) as follows .Ideal resource allocation based on the network load distribution is a main QoS affecting factor. In order to minimize load variance, the proposed self-optimizing model utilizes an intelligent resource allocation pattern. Cost constraints and the upper bound of network power limitation must be considered when distributing resources among entities. In order to allocate power to users over time intervals $t = [1, 2, \dots, T]$, information related to flexibility $\omega_i(t)$ and cost component $c(t)$ must be determined by the big data processing engine through analysis of the received data. $P_i(t)$ denotes the power consumption by user i at moment t and P denotes the restriction on network total power. The problem of real time resource assignment can be formulated as follows:

$$maximize \sum_{k=1}^{K} \alpha_k R_k \quad subject\ to: R_k = \sum_{c=1}^{M} C_k^{(c)} \qquad (8)$$

Replacing $C_k^{(c)}$ with its equivalent in Equation (8) results the formulation of the goal function in Equation (9). The goal of this formula is to maximize the weighted aggregation rate of the network users considering total power constraints. In addition, U and f indicate Utility and Cost functions respectively which are calculated based on network condition and flexibility factor. Also, to achieve optimal solution, variance of allocated resources should be at the minimum possible value.

$$\begin{aligned} maximize \sum_{k=1}^{K}\sum_{c=1}^{M} \Bigg[\alpha_k \rho_k^{(c)} \log\left(1 + \frac{p_k^{(c)} \left|h_k^{(c)}\right|^2}{\Gamma . n_k^{(c)}}\right) + \varphi\, U(P_i(t), \omega_i(t)) \\ - \psi\, f\left(\sum_{i\in N} P_i(t)\right) P_i(t) \Bigg] - \phi \frac{\alpha T}{2} Var\left(\sum_{k\in K} \vec{P}_k\right) \end{aligned} \quad (9)$$

$$bject\ to: \sum_{k=1}^{K}\sum_{c=1}^{M} \alpha_k p_k^{(c)}$$

$$P_{tot}\ \&\ \omega_i(t), \rho_k^{(c)} \ge 0$$

In this model the mentioned coefficients i.e., ω and α are determined based on current user status, service type and network condition through uplink data analysis. Using this method for resource allocation, the self-optimization model satisfies QoS constraints with the minimum cost and low computational load where each user is allowed to transmit at a rate associated with α_k which is related user priority and $\rho_k^{(c)}$ a coefficient with values within the interval 0 and 1 associated with the c^{th} class-carrier for c = 1,2,..,M. The Lagrange equation in Equation (3) is used to take the problem constraints into account.

The channel coefficient for the k^{th} transmitter node on the c^{th} class is denoted by $h_k^{(c)}$ and includes the path loss, Rayleigh fading factor, and constant coefficients for the transmitter and receiver antenna. Moreover, Γ indicates the SINR-gap which is function of coding and modulation and the bit error ratio (BER). For example, for the modulation of non-coded QAM, the SINR-gap Γ will be shown as $\Gamma = {-\ln(5.BER)}/{1.5}$ and $n_k^{(c)}$ denotes the noise of class c for the node k.

$$L(P,\lambda) = \sum_{k=1}^{K}\sum_{c=1}^{M}\left[\alpha_k\rho_k^{(c)}\, log\left(1+\frac{p_k^{(c)}|h_k^{(c)}|^2}{\Gamma.n_k^{(c)}}\right) + \varphi\, U(P_i(t),\omega_i(\mathrm{t})) - \psi\, f\left(\sum_{i\in N} P_i(t)\right) P_i(t)\right] - \phi\,\frac{\alpha T}{2} Var\left(\sum_{k\in K}\vec{P}_k\right) - \lambda\left(\sum_{k=1}^{K}\sum_{c=1}^{M}\rho_k^{(c)}p_k^{(c)} - P_{tot}\right) \quad (10)$$

Assuming that the derivation of Equation (10) is equal to zero, the optimal power allocated to each class-carrier can be obtained by Equation (11). For each iteration of the power assignment algorithm, a current power of the class-carrier that does not satisfy the reliability will be modified and the Lagrange coefficient will be updated accordingly. After R iterations, the algorithm provides an optimal transmission power for each class-carrier as demonstrated by $P_k^{(c)*}$.

$$p_k^{(c)*} = \rho_k^{(c)}\left[\frac{\theta_k}{\lambda} - \frac{\Gamma.n_k^{(c)}}{|h_k^{(c)}|^2}\right]^+ - v - \frac{\alpha T}{2}(\nabla P) \quad (11)$$

Where λ is a Lagrange coefficient associated with transmission power constraints. Equation (4) can be solved by defining auxiliary variable g as:

$$g_k^{(c)} = \frac{|h_k^{(c)}|^2}{n_k^{(c)}}$$

$$P_k^{(c)*} = \rho_k^{(c)}\left[\frac{\theta_k}{\lambda} - \frac{\Gamma}{g_k^{(c)}}\right]^+ - v - \frac{\alpha T}{2}(\nabla P) \quad (12)$$

The optimal power allocation formulation in Equation (12) is very similar to the common waterfilling problem. In this formulation, power is assigned to nodes based on the difference in their weighting coefficients. Each class-carrier of user k is assigned to a flow level equal to $\frac{\alpha_k}{\lambda}$. After waterfilling, the different users have flow levels that are proportional to their weighting coefficients. The users with the higher weighting factors have higher flow levels and can allocate more power to their class-carriers. The Equation (13) can be used to convert a single level waterfilling to a multilevel.

$$\frac{P_k^{(c)*}}{a_k} = \rho_k^{(c)}\left[\frac{1}{\lambda} - \frac{\Gamma}{\theta_k.g_k^{(c)}}\right]^+ - a_k v - (\nabla P)\frac{\alpha.a_k T}{2} \quad (13)$$

Starting with small values for initial Lagrange coefficients, the coefficients are modified in each iteration so that the data rate constraints of different users are satisfied in each node. Each node is treated in turn using the new allocated power until the SINR in the output of the links with temporary failure reach at a minimum level of sensitivity. When the link is recovered, the possibility of obtaining a solution sub-graph in the BD-SON algorithm will increase. In this way, the data rate constraint and dedicated power constraint for all nodes are satisfied and the algorithm will converge.

$$g(\lambda) = \max_{p_k^{(c)}} L(P,\lambda) \approx minimize\ g(\lambda)\,;\lambda \geq 0 \quad (14)$$

The solution to Equation (14) leads to optimal values for maximizing $L(P,\lambda)$. The gradient method can be used to solve the problem to obtain a distributed solution. According to the steepest descent lemma, we have Equation (15):

$$\lambda(t+1) = [\lambda(t) - \gamma.\nabla g(\lambda(t))]^+ \quad (15)$$

where $g > 0$ is the step size and $[Z]^+ = \max\{z,0\}$. Using Equation (8) and according to dual function derivation, we will have Equation (16).

$$\nabla g(\lambda(t)) = \sum_{k=1}^{K}\sum_{c=1}^{M}\rho_k^{(c)}P_k^{(c)*} - P_{tot}(\nabla P)\frac{\alpha T}{2}J \quad (16)$$

Using the Lagrange coefficient (λ), the class-carrier powers in the next iteration can be calculated as in Equation (17):

$$P_k^{(c)*} = \rho_k^{(c)}\left[\frac{\theta_k}{\lambda(t)} - \frac{\Gamma}{g_k^{(c)}}\right]^+ - v - (\nabla P)\frac{\alpha T}{2} = P_k^{(c)}(\lambda(t)) \quad (17)$$

The Lagrange coefficient λ can be updated for the next step in Equation (18) during successive iterations [24].

$$\lambda(t+1) = \left[\lambda(t) - \gamma\left(\sum_{k=1}^{K}\sum_{c=1}^{M}\rho_k^{(c)}p_k^{(c)}(\lambda(t) - P_{tot}\frac{\alpha T}{2}Var\left(\sum_{k\in K}\vec{P}_k\right)\right)\right]^+ \quad (18)$$

Based on the algorithm performance, the SINR can be calculated as the reliability of the transmission in the output of each link. By calculating the value of this parameter, the status of the link becomes clear and the temporary failures can be determined with greater confidence. The new SINR can be used to calculate their new capacities as $C_{ij} = C(SINR_{ij})$.

To achieve transmission reliability, the SINR value at the end point of each link is compared with the minimum SINR required for reliable transmission. In fact, the acceptable SINR threshold is the sensitivity level of the node. If the calculated SINR value is lower than the sensitivity, temporary failure is considered to have occurred for that link and the link will be removed from the set of potential sub-graphs. It is evident that, frequent link failure decreases the possibility of finding a solution sub-graph which satisfies the QoS requirements and increases the cost. To prevent link failure, the power control algorithm is used to gradually increase the transmission power of the node located at the beginning point of the failure links. It should be noted that during all the steps of power control, the total power allocated to all nodes must not exceed a pre-determined value.

6.1 Fixed-Power Efficient Resource Management

Non-linear problem P1 is shown as a mixed integer programming (MIP) framework. The problem now needs to be formulated so that the transmission power is assumed to be fixed; hence, the primal problem P1 will be simplified to problem P2.

$$P2: \max_{x,\varepsilon,G} u(x, \varepsilon, G) \tag{19}$$

s.t. C1, C2, C3, C4, C5, C6

The Simplified primary problem P2 has discrete modality and is a combination of several subproblems. To find an efficient solution for P2, an iterative decomposition approach was applied. Using indicators G and ε, P2 can be modified as:

$$P2.1: \max_{x} U(x) \tag{20}$$

s.t. C1,C2,C4,C5

6.2 Lagrangian Dual Decomposition

Using the goal function, the Lagrangian equation has been formulated to take the problem constraints (P2.1) into account as:

$$L(x,\lambda,\theta) = U(x) - \sum_{j=1}^{N} \lambda_j \left(\bar{\tau}_{min} - \sum_{m=1}^{M+1} x_{jm}\tau_{jm} \right) - \sum_{m=1}^{M+1} \theta_m \left(\sum_{j=1}^{N} x_{jm} P_{jm} - P_m \right) \tag{21}$$

in which λ_j and θ_m are the Lagrange coefficients with positive values. The dual function can be defined as:

$$g(\lambda,\theta) = \begin{cases} \max_{x} L(x,\lambda,\theta) \\ s.t. C2, C5 \end{cases} \tag{22}$$

Using the primal-dual principles, the dual problem can be

Algorithm 1: Efficient UA Algorithm

Step 1: From Subscriber's side
1: **Begin** 2: **if** t = 0
3: **Input:** $\lambda_j(t)$, $\forall j$. Each user equipment evaluates imposed inter-cell interference by checking reference signal of neighbor base stations and the user equipment coose the base station with highest CINR level.
4: **else**
5: j^{th} user equipment finds the levels of μ_{jm} and τ_{jm} through base stations.
6: Selecting the dominant base station (m) based on $m^* = argmax_m(\mu_{jm})$
7: Update $\lambda_j(t)$ based on (16).
8: **end if**
9: $t \leftarrow t + 1$.
10: The request of the user equipment will be feedback towards selected base station and degree of $\lambda_j(t)$ will be broadcast accordingly.
Step 2: From Base Station Perspective 1:**Begin**
2: **if** t = 0
3: Input $\theta_m(t), \forall m$.
4: **else**
5: Consider the matrix x as the uptodate user association indicator
6: Updating $\theta_m(t)$ exactly based on (17), respectively.
7: All base stations compute μ_{jm} and τ_{jm}according to Non-ortogonal multiple access.
8: **end if**
9: $t \leftarrow t + 1$.
10: All base stations multicast the degree of μ_{jm} and τ_{jm}.

used to simply produce the solution. Using P2.1, the dual problem can be defined as shown in Eq. (13) as:

$$\min_{\lambda,\theta} g(\lambda,\theta) \tag{23}$$

Using the dual Lagrangian coefficients λ_j and θ_m, the optimum value of parameter x for the Lagrange formula can be calculated as:

where, in Eq. (14), $m^* = \mathrm{argmax}_m(\mu_{jm})$with:

$$x_{jm}^* = \begin{cases} 1, & \text{if } m = m^* \\ 0, & \text{otherwise} \end{cases}, \tag{24}$$

The optimal solution in Equation (24) means that subscribers choose base stations (which supply the maximum achievable throughput) according to the grid power consumption. Because the goal of the dual function is not differentiable, the subgradient method was applied to achieve the solution (λ^*, θ^*) with maximum optimality of the dual problem. This can be expressed as:

$$\lambda_j(t+1) = [\lambda_j(t) - \delta(t)\left(\sum_{m=1}^{M+1} x_{jm}\tau_{jm} - \bar{\tau}_{min} \right)]^+, \tag{25}$$

$$\theta_m(t+1) = [\theta_m(t) - \delta(t)\left(P_m - \sum_{j=1}^{N} x_{jm}P_{jm}\right)]^+, \quad (26)$$

In Equations (25) and (26), $[a]^+ = \max\{a, 0\}$, t is the indicator of iteration and $\delta(t)$ denotes the step size. Of the various step size patterns available, such as fixed and decreasing step sizes, in this study, decreasing step size was applied as introduced by Zhang [25]. After achieving the optimal values (λ^*, θ^*) using Eqs. (25) and (26), the optimal value of index x will be the response to main problem P2:1. Using the iterative decomposition method, user association can be designed in both centralized and decentralized modes, which this a specific capability of the proposed approach. The centralized user association system needs an intensive controller which has access to global channel state information and is aware that the subscriber is served by a specific base station. This article suggests a decentralized user association approach in which there is no requirement for centralized control data using algorithm 1. Because the convergence requirements (described by in [25]) are satisfied through the proposed algorithm, convergence is definite.

For each iteration, the proposed algorithm has complexity order $O((M+1)N)$. Convergence of the resource management algorithm appears to be fast because convergence occurs in less than 40 repetitions. This is significantly less than the Brute Force approach with complexity order $O((M+1)^N)$. The broadcast operations applied in the proposed model have little effect on the order of time complexity.

6.3 Genetic-based UA Algorithm

Next, a genetic algorithm (GA) based on user association is proposed to solve main problem P2.1. The aim is to compare the results of proposed algorithm 1 with a genetic-based algorithm. As stated by Abdelaziz [30], a genetic-based algorithm has good functionality if the number of potential solutions is adequate. In particular, each chromosome represents a feasible solution which meets the requirements of problem P2.1 and is formulated as:

$$D_i = \{[m_{1i}], [m_{2i}], \dots, [m_{Ni}]\}, i \in \{1, \dots, K\}, \quad (27)$$

In Equation (27), m_{ij} denotes the gene for the base station to which user equipment j is connected, with a variable amount between interval [1 M+1]. K denotes the density indicator of the population. In any production process, all chromosomes will be assessed based on compatibility to allows selection of the most compatible chromosomes and generates the most compatible children. In the goal of primal problem P2.1, the compatibility of chromosome D_i is computed as:

In this evaluation, all chromosomes having rank r are ranked according to their fitness and compatibility. For a chromosome, the possibility of producing child is equal to $\rho_s(r) = \frac{q(1-q)^{r-1}}{1-(1-q)^k}$ with predefined value q that has been defined by in [38]. During all production stages, a monotonous crossover action with possibility ρ_c is applied for generation of a child by exchanging or merging genes using the generator chromosomes. A monotonous mutation action with probability ρ_m has been applied. The generation method is repeated until the highest amount of production is achieved as shown in algorithm 2.

Considering the highest amount of production and a constant population density equal to K, the computational complexity order of the proposed framework is equal to $O(\Omega K\log(K))$ [39].

The performance of the user association based on genetic-based algorithm depends upon the number of generations and the size of the population [38]. The simulation results confirm that the effectiveness of algorithm 1 is greater than genetic-based algorithm 2. When the density of the population in the genetic algorithm is not sufficient, it has lower computational complexity.

Algorithm 2: Genetic-based UA

1: **if** t = 0
2: **Input** set of possible chromosomes $\{D_i\}$ by population density equal to K and the topmost number of production t_{max}.
3: **else**
4: Ordering $\{D_i\}$ according to the degree of compatibility roaming in (19).
5: According to the selection possibility $\rho_s(r)$, chromosomes have been selected to generate offspring by identical jumping operations.
6: **If** in case of exceeding the upper bound of generation
7: Consider $x_{jm}^* = \{D_i^*\}$, where $\{D_i^*\}$ indicates the possible chromosome with maximum degree of compatibility
8: **break**
9: **else**
10: $t \leftarrow t+1$.
11: **end if**
12: **end if**

The aforesaid approach obtains an optimal user association framework for P2.1. By achieving the optimal value of user association $X = [x_{jm}^*]$, parameters (G, ε) are achievable according to linear-based programming (LP) method in Eq. (28):

$$P2.2: \min_{\varepsilon, G} \sum_{m=1}^{M+1} G_m \quad (28)$$

s.t. C3.C6.

P2:2 is effectively solvable by applying software tools, such as CVX tool.

If power sharing is not permitted, as for $\varepsilon_{mm'} = 0, \forall m$, the optimal amount of grid requested power (G) is for P2.2 based on user association solution $X = [x^*_{jm}]$. This can be calculated using Eq. (29) as:

$$G^*_m = [P_m - E_m]^+, \quad (29)$$

where $P_m = \sum_{j=1}^{N} x^*_{jm} P_{jm}$.

Using the solutions obtained for sub-problems P2.1 and P2.2, algorithm 3 is introduced as an optimal framework for solving P2 iteratively.

7. Result

7.1 Mistakes Amount of Gathered & Analyzed Data

The performance of our proposed approach was evaluated under different scenarios having specific assumptions to simplify. For this purpose, BD-SON algorithm was compared with current self-optimization solutions and SOTA schemes with non-self-organization framework. SOTA refers to previous schemes to find efficient and optimal solution by various approaches. Some of the main SOTA approaches mentioned in the reference part. Also, we have compared the proposed approach with most compatible SOTA to the described model.

In these scenarios, channel Drop Rate (DR), Downlink Quality and Throughput were applied as target indicators.

It should be mentioned that we applied some various software planes to assess the proposed approach. For example, pre-process part of the model is done by MATLAB and we achieved KPI results by some network planning tools which they are able to monitor networks as real time. Also, some parts of the network modeling like as network configuration definition is done bvia network simulator software.

The volume of data required by the BD-SON is significantly higher than for the SOTA. Moreover, the volume of data stored has a significant relationship with the density of eNBs and UEs. According to Figure 4, this value for SOTA is relatively constant based on the problem assumption in which the volume of the data gathered by SOTA is somewhat related to the density of the UEs, not the density of the eNBs. It should be noted that the processing speed of BD-SON is directly related to the volume of the data stored in the register and the computational complexity of data frame decoding.

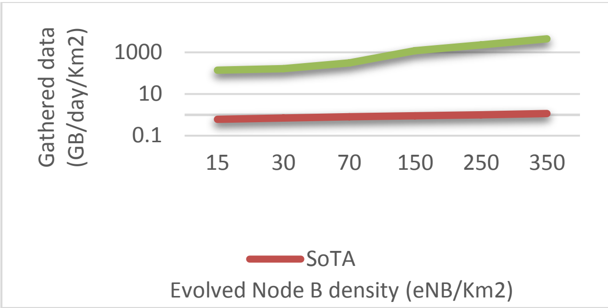

Figure 4. Amount of gathered and analysed data for decision-making in BD-SON with comparison to SOTA

7.2 Channel Quality Estimation:

Our introduced approach not only reduce the interference of its specified cell, but also increase the network functionality by alleviating interference to adjacent cell areas. Figure 5 demonstrates the coverage probability considering SINR values based on BD-SON functionality in comparison with Q-Learning SON scheme [24] and the SOTA schemes with non-self-organization feature.

We have evaluated the performance of the network in a cluster having 15 sites in the presence and the absence of the SON engine. As shown in Figure 5, the level of DCR and the signalling drop rate (SD_Drop_Rate) have significantly degradation after self-optimization model activation.

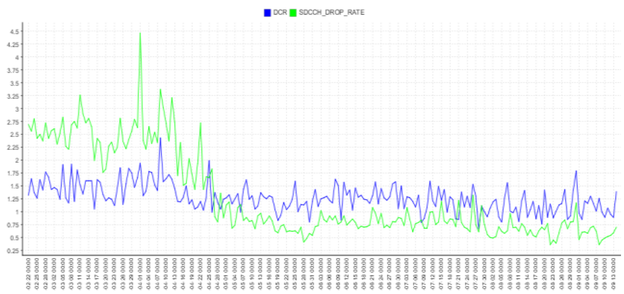

Figure 5. Decreasing the Drop Rate, by enabling the self-optimization model

The vertical axis of the plot indicates the percent of Drop Call Rate and the horizontal axis denotes the Decreasing the Drop Rate, by enabling the self-optimization model

Downlink quality will also increase when the SON model is enabled. As shown in Figure 6, when the self-optimization scheme was applied in the BSS layer, the user quality of service is stable without any considerable fluctuation.

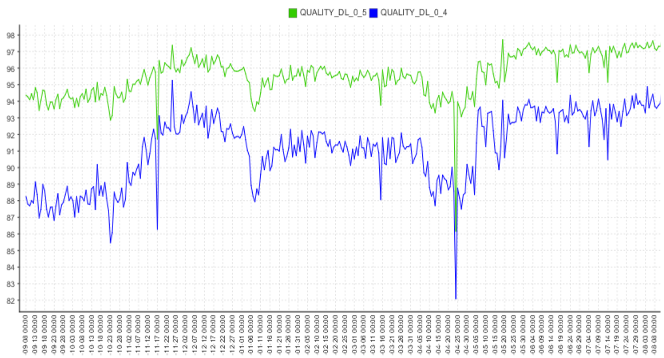

Figure 6. Stable downlink quality, by enabling the self-optimization model

7.3 Timing Complexity:

The vertical axis of the plot indicates the percent of Quality DL/UL and the horizontal axis shows the Stable downlink quality, by enabling the self-optimization model

Then, by evaluating the functional timing of the proposed model under three distinct modes in a general manner, solely based on learning, and self-optimization based on learning and recommender system, we tried to evaluate the effects of using state space reduction technique on reducing the time required for the decision-making process.

As the outputs in figure 7 show, in proportion to the lower number of network parameters, the decision-making time is solely dependent on learning algorithms, and using the recommender system reduces the required time. However, by increasing the number of parameters, statistical analyses effectively reduce the number of parameters influencing the target indicators and the decision-making process becomes simpler and quicker. The obtained outputs indicate that in the small state space, using feature extraction and parameter selection modules will not have a significant impact on the performance of the self-optimization engine, and it will increase the volume of computations and, in turn, the response time. On the other hand, by increasing the number of network parameters, the presence of these modules will effectively reduce the time required for the decision-making process.

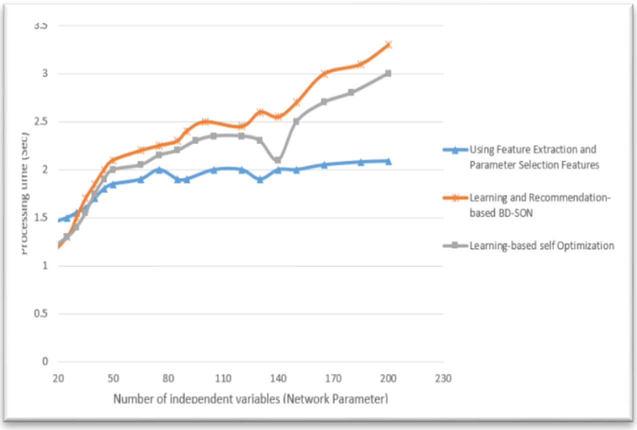

Fig.7 The effect of engine’s component on timing complexity

Moreover, using the recommender system, along with the learning algorithm based on support vector, will add an approximately constant time to the processing duration. This constant value is related to the passing of the components of the set of commands through the rule-based filter. If we use inference engine to increase the capability of the recommender system, this time difference will obviously increase.

7.4 Accuracy Evaluation:

In this section of assessing the self-optimization model, we try to evaluate the accuracy and recall of the learning section, which is used for predicting the target efficacy indicators.

As mentioned earlier, the learning tool used in this study is the support vector machine. However, the reason behind selecting this tool as the main unit for estimating the status of the efficacy measures is the higher level of accuracy and recall than other learning methods. As can be seen from the figure 8, the value of the ROC indicator is somewhat higher for the learning method based on support vector compared to other learning methods. Therefore, using this learning method will allow for reaching decisions with a higher level of accuracy. The ROC indicator used is related to the ratio of correct true positive decisions to incorrect true positive decisions, which is a valuable criterion for determining the accuracy level of the network’s decision-making. However, if we want to use a more powerful criterion, we can use the F-measure.

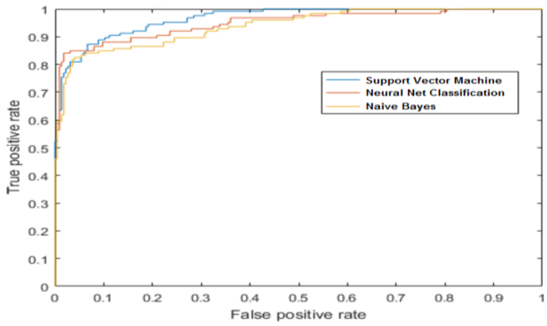

Fig.8. SVM accuracy in comparison with other ML methods

As can be seen from the figure below, support vector-based learning method creates a bigger area under the curve than the other two learning methods do, thus indicating the higher accuracy level of decision-making in this method for this type of data. Then, we evaluate the performance of these three machine learning methods in the real-life mobile network data platform. In this scenario, these three machine learning methods are analysed using three important criteria—precision, recall, and F-measure. The obtained results confirm the findings of previous schemes, presented in the following figure.

As can be seen from the table 1, besides precision and recall factors, the highly accurate F-measure is used also for assessing the performance accuracy of the three learning methods, and the results indicate the superiority of the support vector learning method over the other two methods based on the detection accuracy level.

Table 1. SVM accuracy level in comparison other ML methods

	Precision (%)	Recall (%)	F-measure (%)
SVM	91.04	87.16	89.05
Naive Bayes	89.46	87.79	88.61
Neural Net Classification	79.64	81.27	80.44

It should be noted that Presicion, Recall and F-Measure factors are accessible via following formulas:

$$\text{Precision} = \frac{t_p}{t_p + f_p}$$

$$\text{Recall} = \frac{t_p}{t_p + f_n}$$

$$F_{\text{measure}} = 2.\frac{\text{Precision. Recall}}{\text{Precision} + \text{Recall}}$$

Which these factor's components are defined as below:

Recall ~ Sensitivity

Precision ~ Positive Predictive Value (PPV)

t_p: True Positive Rate

f_p: False Positive Rate

f_n: False Negative Rate

t_n: True Negative Rate (accuracy)

Based on the assessments, the support vector learning method is used for predicting the target efficacy indicator in the monitoring and learning modules.

In this scenario, the feature extraction and parameter selection are evaluated as the basic modules in the process of data analysis. In order to evaluate the results of the proposed scheme regarding the network's intelligence in selecting parameters influencing the target indicator, we want to analyse the output from implementing the proposed method in a cluster involving five basic stations and 15 cells under KPI mode and drive test mode. In this scenario, the throughput index, as the target efficacy indicator, is below expectations. Therefore, it should be determined what kind of solution based on the proposed self-optimization model can be utilized to improve this indicator. To begin with, we consider the feature extraction component. In this section, the system identified the parameters whose values were directly related to the throughput indicator. In this section, about 50 parameters with satisfactory significance level are selected. Out of the selected indicators, transmission power, transmission mode, neighbourhood threshold, and various timing factors are related to the 'Idle' and 'Dedicated' modes.

KPI control:

In the next stage, by carrying out a linear multiple regression model on the current state space, the level of relations and the effects of the extracted parameters on the values of the throughput indicator were identified. The relations between some of the parameters and the target indicator were logarithmic, and some were exponential. After identifying the impact coefficients, about 10 main parameters were selected from among all the parameters

based on the obtained impact factor. It is interesting that, several parameters, selected in the feature extraction phase with a high correlation coefficient, obtained a low impact factor during the assessment of the effects of parameters on the throughput measure and were thus eliminated. We can particularly mention the transmission mode indicator and the physical characteristic of antenna E-tilt, which generally have a direct impact on the level of throughput. The reason for this can be related to the unstable state of the network, as well as the effects of the network parameters on the impact of a certain network parameter on the efficacy indicator. In the statistical tests, only the correlation of the parameters and the efficacy indicator is measured. In the regression analysis, the effects of each parameter on the general state and the presence of other network parameters on the target indicator will be evaluated.

In the support vector-based learning unit, the selected parameters are used as inputs. In this phase, the predicted throughput level is considered, which was lower than the threshold value. Therefore, the decision-making operation is continued by determining a set of specific commands.

While the number of handover requests has significantly come down, the handover success rate (HOSR) has only reduced as much as a few hundredth of a per cent. As can be seen from the figure 9, this reduction will not have a negative impact on the network's efficiency. In fact, a 0.5% reduction in the handover success rate is the negative impact of the decision made by the self-optimization engine for increasing the throughput. However, this degradation is not high enough to get the efficiency indicator of handover success rate out of the acceptable range.

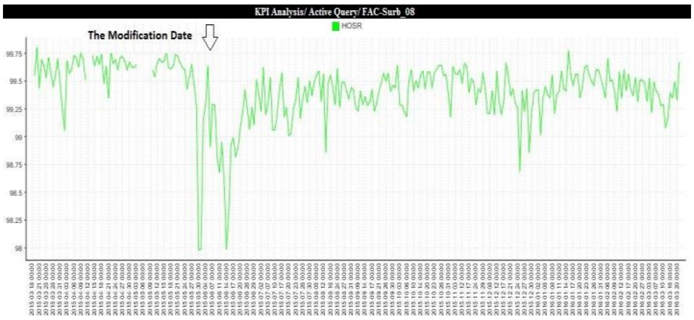

Figure 9. HO success rate, by enabling the self-optimization model

Figure 9 which is relevant to Stable downlink quality, by enabling the self-optimization model, the vertical axis of the plot indicates the percent of Handover Success rate and the horizontal axis shows the timing of the algorithm execution.

On the other hand, the main impact of tightening the handover conditions is a reduction in the traffic load of the network's signalling. The figure 10 shows that, after applying this change in the network, the traffic load of network's signalling is degraded 30%. This can be very useful in reaching an optimal network. Under these conditions, unnecessary network signalling resources can be allocated to the data transmission traffic channels.

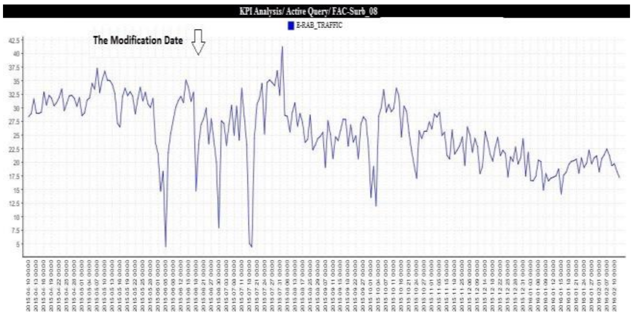

Figure 10. Dercreasing of signaling rate, by enabling the self-optimization model

Figure 10 which is relevant to Dercreasing of signaling rate, by enabling the self-optimization model, the vertical axis of the plot indicates the amount of the traffic and the horizontal axis shows the timing of the algorithm execution.

One of the issues that should be considered is that, by reducing the control signals, the conditions of the network regarding service retainability and service quality must not be changed. One of the effective measures in providing services with SLA is the quality level of the communication channels from the viewpoints of uplink and downlink. In this scenario, by changing the ranges of the predefined network parameters, we will look at the levels of these measures with respect to the quantitative aspects as well as stability.

As can be seen from the figure 11, there is no negative change in the quality of uplink and downlink links. However, this is also related to the reduction in the number of clients. Moreover, the stability in the quality level of uplink radio frequency channels is one of the results of applying the decisions of the proposed self-optimization engine, which has a positive impact on the synchronous communication between UE and the network core through Backhaul.

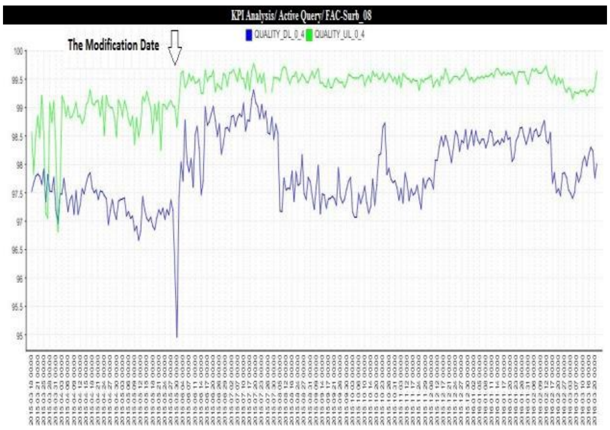

Figure 11. Quality stability after enabling the self-optimization model

Figure 11 which is relevant to horizontal axis shows the

Quality stability after enabling the self-optimization model, the vertical axis of the plot indicates the Quality and the horizontal axis shows the timing of the algorithm execution.

However, by assessing all the important efficiency indicators influencing the network, we tried to quantitatively analyse the reaction of the network status to the decisions of the self-optimization engine. As has been seen so far, the decisions for increasing the level of the network's throughput rate do not have significant negative impacts on the other efficiency indicators of the network, and the positive effects of introducing the recommender module in the structure of the self-optimization engine are obvious. In the following, the effects of the output of the self-optimization model on the target indicator of throughput rate are discussed and evaluated.

As can be seen from the figure 12, by allocating radio signalling resources to the traffic channels of data transmission, the data transmission rate, particularly in the uplink direction, is significantly increased. Of course, the reduction in the number of clients is also important in this regard. However, this increase in the rate and in the level of traffic resources is obvious. Allocating a larger number of resource blocks to traffic connections will lead to a service with a higher throughput rate and an average 20% increase is clearly seen.

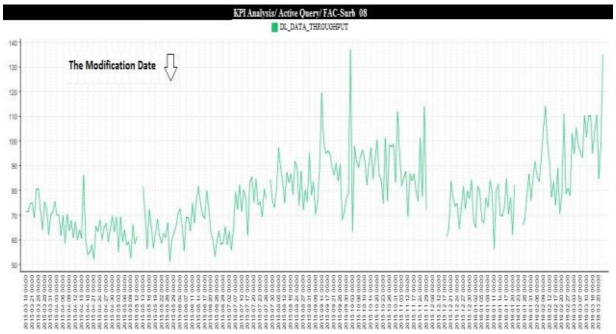

Figure 12. Throughput enhancement, by enabling the self-optimization model

Figure 12 which is relevant to Throughput enhancement, by enabling the self-optimization model, the vertical axis of the plot indicates the throughput level and the horizontal axis shows the timing of the algorithm execution.

Therefore, it can be concluded that the performance of the self-optimization engine in this scenario is highly satisfactory, and that using this model, the system will be able to change the parametric structure of the network for a target indicator in a way that the value of this indicator is increased to a satisfactory level without any significant negative impact on the other indicators. It is worth mentioning at this point that the basis for the functioning of the proposed self-optimization model is iterative. The value of the target indicator after the change is observed and based on the new value of the indicator after the changes are applied to the network parameters in each cycle, the decision-making process is repeated so that the target indicator's value reaches a satisfactory level.

In this section, we tried to evaluate the changes in the levels of the network's efficiency indicators by implementing the self-optimization model. It can be concluded, that under normal conditions, the predictions for the performance of the model accurately depict the correct functioning of the model. However, in order to prove the positive impact on the model, it should be assessed in larger-scale networks as well. In the next section, we try to evaluate the performance of the proposed model with similar well-known schemes.

7.5 Coverage Assessment:

Our introduced approach not only reduce the interference of its specified cell, but also increase the network functionality by alleviating interference to adjacent cell areas. Figure 13 demonstrates the coverage probability considering SINR values based on BD-SON functionality in comparison with Q-Learning SON scheme [24] and the SOTA schemes with non-self-organization feature.

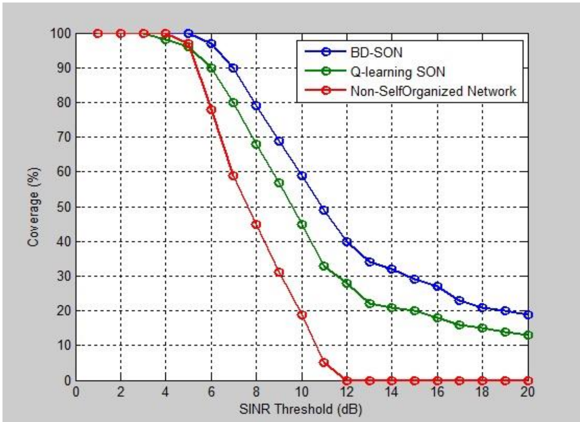

Figure 13. Coverage probability of Q-Learning SON, BD-SON and Non-Self Organized network based on needed SINR

Figure 14 shows the effects of the proposed model on throughput based on number of layers defined for the network. As expected, the self-optimization model using big data increased the throughput. Also, the network capacity has been approached to the maximum theoretical capacity. Using this model, the SINR indicator for macro users was maintained at an ideal level and the throughput of each cell was improved by increasing the number of layers.

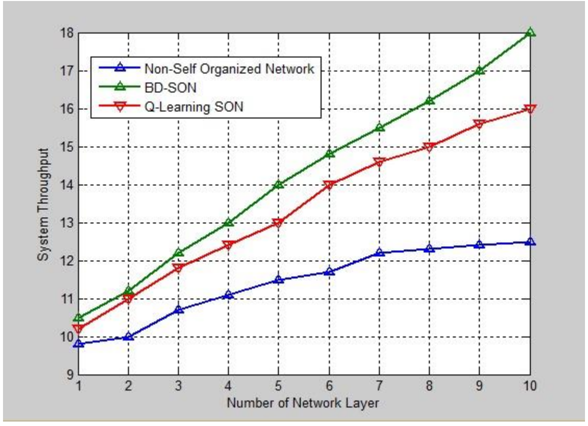

Figure 14. Increasing throughput by increasing the number of layers in BD-SON in comparison with Q-Learning

8. Conclusion and Future Works

This study introduced self-optimization technique as a powerful tool for distributed network management, consistent with next-generation wireless networks. In proposed method, coded flow multicasting has been utilized for guaranteeing the QoS constraints by finding an optimal sub-graph in a flow-based optimization framework. the load balanced gradient power allocation (L-GPA) algorithm was also applied for the QoS-aware multicast model to accommodate the effect of transmission power level based on load distribution to increase link resistance to temporal failure caused by interference and noise. In comparison to other schemes proposed method can better satisfy the QoS requirements of multicast sessions. The simulation results prove that using introduced approach considerably increases the chance of finding an optimal sub-graph. Also, Experimental results show that the proposed method can improve the throughput and quality of service in next-generation wireless networks. A further study with more focus on other requirements of next generation wireless networks are therefore recommended. Also, further research might investigate determining the self-healing intelligently based on environmental conditions. From KPI perspective, experimental results prove that despite the tremendous volume of the analyzed data – which is hundreds of times bigger than usual methods – it is possible to improve the KPIs, such as throughput, up to 30 percent by optimal resource allocation and reducing the signaling load. Also, uplink quality improvement and 15-percent increment of the coverage area under satisfied SINR conditions can be considered as outcome of the proposed scheme. Also, computational complexity analysis will be one of the appropriate ideas as the future works of the scheme.

References

[1] A. Imran, A. Zoha and A. Abu-Dayya, “Challenges in 5G: how to empower SON with big data for enabling 5G,” IEEE Network, Vol. 28, No. 2, 2014, pp. 27–33.

[2] N. Baldo, L. Giupponi and J. Mangues-Bafalluy, "Big Data Empowered Self Organized Networks," European Wireless 2014; 20th European Wireless Conference, 2014, pp. 1-8.

[3] R. Murugeswari, S. Radhakrishnan, and D. Devaraj, “A multi-objective evolutionary algorithm based QoS routing in wireless mesh networks,” Applied Soft Computing, Vol. 40, No C, 2016, pp. 517–525.

[4] Jun, Ersin Uzun, and Jose J. Garcia-luna-aceves. "NETWORK CODING FOR CONTENT-CENTRIC NETWORK." U.S. Patent 20,160,065,685, issued March 3, 2016..

[5] H. Zhang, C. Jiang, R. Q. Hu and Y. Qian, "Self-organization in disaster-resilient heterogeneous small cell networks," IEEE Network, vol. 30, No. 2, 2016, pp. 116-121.

[6] S. Berger, M. Simsek, A. Fehske, P. Zanier, I. Viering, and G. Fettweis, “Joint Downlink and Uplink Tilt-Based Self-Organization of Coverage and Capacity Under Sparse System Knowledge,” IEEE Transactions on Vehicular Technology, Vol. 65, No. 4, 2016, pp. 2259–2273.

[7] M. Huang and XU. Zhang, “Enhanced automatic neighbor relation function for 5G cellular systems with massive MIMO,” 2017 IEEE International Conference on Communications (ICC), 2017, PP 1-6. 2017.

[8] K.-L. Yap and Y.-W. Chong, “Software-Defined Networking Techniques to Improve Mobile Network Connectivity: Technical Review,” IETE Technical Review, 2017, pp.1-13.

[9] Gebremichail and C. Beard, “Fade Duration Based Neighbor Cell List Optimization for Handover in Femtocell Networks,” International Journal of Interdisciplinary Telecommunications and Networking, Vol. 9, No. 2, 2017, pp. 1–15.

[10] T. Han and N. Ansari, “Network Utility Aware Traffic Load Balancing in Backhaul-Constrained Cache-Enabled Small Cell Networks with Hybrid Power Supplies,” IEEE Transactions on Mobile Computing, Vol. 16, No. 10, 2017, pp. 2819–2832.

[11] Guita, José, Luis M. Correia, and Marco Serrazina. "Balancing the Load in LTE Urban Networks via Inter-Frequency Handovers." M.S. thesis, Instituto Superior Técnico ,University of Lisbon, Lisbon, Portugal, 2016.

[12] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, “Big data-driven optimization for mobile networks toward 5G,” IEEE Network, Vol. 30, No. 1, 2016, pp. 44–51.

[13] S, L, B. Iwan, R. Nicolas, Q. Ripault, J.R. Andrade, S. Han, H. Kim, "Self-optimization of plasmonic nanoantennas in strong femtosecond fields." Optica, Vol.4, No.9, 2017, pp. 1038-1043.

[14] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, “Toward Massive Machine Type Cellular Communications,” IEEE Wireless Communications, Vol. 24, No. 1, 2017, pp. 120–128,.

[15] Gillot, David, and John Yue Jun Jiang. "Method and system for providing roaming intelligence (RI) to a host network operator for its roaming traffic." U.S. Patent 9,338,663, 2016.

[16] S. Fong, C. Fang, N. Tian, R. Wong, B. W. Yap, "Self-Adaptive Parameters Optimization for Incremental Classification in Big Data Using Neural Network." In Big Data Applications and Use Cases Springer International Publishing, 2016.

[17] Agrawal, Himanshu, and Krishna Asawa. "New Architecture for Dynamic Spectrum Allocation in Cognitive Heterogeneous Network using Self Organizing Map." arXiv preprint arXiv, 2016.

[18] G. Foster, S. Vahid, and R. Tafazolli, "SON Evolution for 5G Mobile Networks," Fundamentals of 5G Mobile Networks, USA, Wiley, 2015.

[19] H. Goudarzi and M. Pedram, "Hierarchical SLA-Driven Resource Management for Peak Power-Aware and Energy-Efficient Operation of a Cloud Datacenter," IEEE Transactions on Cloud Computing, Vol. 4, No. 2, 2016, pp. 222–236.

[20] C. Segura, C. Coello, G. Valladares, C. Leon, " Using multi-objective evolutionary algorithms for single-objective constrained and unconstrained optimization." Annals of Operations Research, Vol.240, No.1, 2016, pp. 217-250.

[21] Z. M. Fadlullah, D. M. Quan, N. Kato, and I. Stojmenovic, "GTES: An Optimized Game-Theoretic Demand-Side Management Scheme for Smart Grid," IEEE Systems Journal, Vol. 8, No. 2, 2014, pp. 588–597,.

[22] Y. Liu, C. Yuen, S. Huang, N. Ul Hassan, X. Wang and S. Xie, "Peak-to-Average Ratio Constrained Demand-Side Management With Consumer's Preference in Residential Smart Grid," IEEE Journal of Selected Topics in Signal Processing, Vol. 8, No. 6, 2014, pp. 1084-1097.

[23] A. De Waegenaere, J. L. Wielhouwer, " A breakpoint search approach for convex resource allocation problems with bounded variables."Optimization Letters ,Vol.6, No.4, 2012, pp. 629-640.

[24] A. Galindo-Serrano, "Self-organized Femto-cells: A Time Difference Learning Approach," Ph.D. thesis, Universitat Politecnica de Catalunya (UPC), Barcelona, Spain, 2013.

[25] Y. Jiang, Q. Liu, F. Zheng, X. Gao and X. You, Energy-efficient joint resource allocation and power control for communications, IEEE Transactions on Vehicular Technology 65(8) (2016), 6119–6127

Abbas Mirzaei received the B.S. degree in Computer Engineering from Azad University, Gazvin Branch, Iran in 2006, and M.S. degree in Computer Engineering from Azad University, Tabriz Branch, Iran, in 2008, and Ph.D. degree in Computer Engineering from Malek Ashtar University of Technology, Tehran, Iran in 2018. He currently works as an assistant professor at the Department of Computer Engineering, Islamic Azad University, Ardabil Branch. His research interests include Computer security and reliability, Artificial intelligence, Computer communications (networks), Cellular network optimization, Heterogeneous networks, Data mining, and Wireless resource allocation.

Amir Rahimi received the B.S. degree in Computer Engineering from Sabalan University, Ardabil, Iran in 2015 and M.S. degree Computer Engineering from Moghadas Ardabil University, Ardabil, Iran, in 2019. His area research interests include Artificial intelligence, Cellular network optimization, and Data mining.