Fuzzy Q-Learning for Mobility Robustness Optimization in Wireless Networks

Andreas Klein, Nandish P. Kuruvatti, Jörg Schneider, Hans D. Schotten
Institute for Wireless Communications and Navigation
University of Kaiserslautern
Email: {aklein,kuruvatti,schneider,schotten}@eit.uni-kl.de

***Abstract*—The high popularity of smartphones and mobile PCs is expected to increase wireless data traffic in the order of 1000 times by 2020 [1]. However, the current situation of Mobile Network Operator (MNO)s is characterized by increasing margin pressure due to declining revenues and an increasing cost base. Self-optimization functionalities, e.g. for Mobility Robustness Optimization (MRO), are essential means for reducing Operational Expenditure (OPEX). In particular, mobile user groups or moving networks at high speeds impose challenges and may severely degrade network performance as well as user experience. The Fuzzy Q-Learning-based approach presented in this paper aims at providing a generic basis for enabling self-optimizing and self-healing network operations. The designed concept consists of the following key components: Fuzzy Inference System (FIS), heuristic Exploration/Exploitation Policy (EEP), and Q-Learning (QL). Its performance in a reference scenario is compared with a trend-based handover (HO) optimization scheme presented in [2] and a scheme that assigns time-to-trigger (TTT) values based on velocity estimates.**

I. Introduction

Arising and spreading 4G Radio Access Technology (RAT) deployments promise to provide significant increases in capacity and enhanced Quality of Service (QoS) for mobile broadband applications, such as multimedia, video conferencing, etc. However, the current situation of Mobile Network Operator (MNO)s is characterized by increasing margin pressure due to declining revenues and an increasing cost base. In particular, OPEX have become a more significant part of operators' cost structure. The major challenge for MNOs is to provide mobile broadband and potentially new services in a high-quality, but cost-effective manner. Thus, there is a need for flexible and cost-efficient network management and for supporting a variety of services ranging from real-time to high data rate on demand services. One important feature that may particularly allow MNOs to stand out from their competitors are learning-based approaches for self-optimization of mobility support in high speed scenarios with recurring traffic conditions as well as self-healing in case of hardware failures. In 2007, the Next Generation Mobile Networks (NGMN) alliance released Self-Optimizing Network (SON) use cases [3] and recommendations on SON requirements [4]. Currently rolled out LTE networks are required to support a wide range of user velocities ranging from low and nomadic to even high speed mobility at up to 350 km/h. Since changes in user mobility directly affect network performance with respect to mobility-related KPIs, it is important to develop self-optimization schemes that are able to adaptively adjust HO parameter settings given varying or recurring mobility scenarios. Recently, many SON approaches have been presented that aim at enhancing mobility robustness in LTE networks. In [5], a self-optimizing algorithm for HO parameter control is presented that iteratively adjusts the parameter settings of time-to-trigger (TTT) and HO margin (HOM) depending on the oscillations experienced between different cell pairs. A first trend-based HO optimization algorithm, further referred to as TBHOA, is presented in [2], where the MNO defines a set of major KPIs and assigns certain weights that reflect the importance of these KPIs with respect to the MNO's policy. Whenever a specific KPI is found to be below the MNO-defined performance target threshold for a predefined amount of time, the TBHOA tries to optimize system performance by executing actions stored in a look-up table (LUT). The distributed HO optimization procedure presented in [6] combines target and source cell information, such as radio link failure (RLF) indication messages, HO reports, and detected ping-pong HOs, into KPIs indicating too late or too early HOs and determines, based on a weighting applied to these KPIs, whether the values for A3-Offset, TTT, or CIO for considered neighbor cell are modified towards an earlier or later trigger for HO to the neighbor cells. Of particular concern are the HOs of fast moving UEs, since they experience higher HO rates and potentially suffer higher failure rates than slow moving UEs. [7] proposed a HO parameter optimization algorithm that detects the change in UE mobility through the change in HO failure events and adaptively adjusts HOM to the UE mobility. It is demonstrated that the algorithm effectively reduces both the HO failure and ping-pong HO rates, while the trade-off between the two rates taken into consideration changes in several mobility scenarios. [8] studied network-wide, cell-specific, and cell-pair specific optimization approaches for intra-frequency MRO and presented heuristic solutions for adapting *hysteresis* (HYS) values, where the number of RLFs and ping-pong HOs is incorporated via cost functions and significant gains can be achieved the more location-specific the parameters are defined. The authors of [9] exploit vehicle context information, e.g. position and trajectory, to adaptively optimize street-specific HO parameters. In contrast to the majority of the afore mentioned approaches, the FQL-based approach presented in this paper allows for autonomously adapting HO parameters to local, cell sector-specific conditions as well as setting HO parameters in a cell pair-specific manner, while simultaneously accounting for several KPIs, such as dropping, HO failure, and ping-pong HO ratios. The remainder of this paper is organized as follows: section II describes HO procedure in LTE networks and introduces mobility-related KPIs, section III illustrates the developed Fuzzy Q-Learning-based MRO scheme, section IV

presents simulation results, and section V concludes the paper.

II. Handover Event and KPIs

A. Handover Event

For intra-frequency mobility in cellular LTE networks, HO processes a triggered as soon as the condition for the so-called A3 event [10] measurement reporting becomes true. The following simplified HO condition has to be satisfied for a certain amount of successive time instances, referred to as *time-to-trigger* (TTT), until reporting is triggered:

$$\begin{aligned} RSRP_t + CIO_t &> HYS + CIO_s + RSRP_s, \quad (1) \\ RSRP_t - RSRP_s &> HYS + (CIO_s - CIO_t), \\ RSRP_t - RSRP_s &> HOM(s,t), \end{aligned}$$

where $RSRP_t$ and $RSRP_s$ denote the Reference Signal Received Power (RSRP) values and CIO_t and CIO_s denote the CIO values of the target and serving eNBs, respectively. Solving equation 1 for $(RSRP_t - RSRP_s)$ yields a relation between the $RSRP$ difference and the overall HOM with respect to *serving* eNB s and *target* eNB t. Changes of the HYS, TTT, and the CIO values have an impact on the quality during HOs, on the success rate of HOs, on the occurrence of unwanted ping-pong HOs, and on the load distribution between different cells.

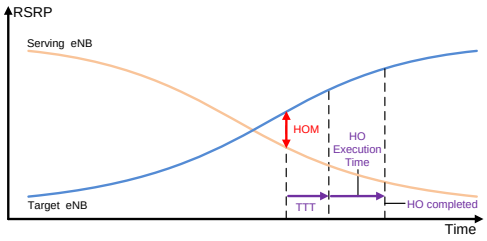

Fig. 1. Illustration of Handover Event and Relevant Parameters

B. KPIs

For assessing HO performance and optimizing mobility support several KPIs, such as *connection dropping ratio* (CDR), *HO failure ratio* (HFR), and *ping-pong HO ratio* (PHR), are evaluated:

$$CDR = \frac{N_{dropped}}{N_{connected}}, \quad (2)$$

where $N_{dropped}$ denotes the number of dropped connections and $N_{connected}$ the overall number of connections during observation time window T, respectively. The HO failure ratio is calculated as follows:

$$HFR = \frac{N_{HO\_fail}}{N_{HO\_fail} + N_{HO\_success}}, \quad (3)$$

where N_{HO_fail} and $N_{HO\_success}$ denote the number of failed and successful HOs, respectively, and altogether yield the overall number of HOs.

$$PHR = \frac{N_{HO\_ping-pong}}{N_{HO\_success}} \quad (4)$$

represents the proportion of monitored ping-pong HOs with respect to the number of successful HOs.

III. Proposed Fuzzy Q-Learning-based Optimization Scheme

The reinforcement learning scheme used by the developed approach faces the issue of exploring the whole *universe of discourse* for attaining a rich body of experience and learning the best action for any situation. In order to reduce training complexity and efforts, the Fuzzy Q-Learning (FQL)-based scheme presented in this paper applies a limited EEP and first classifies the universe of discourse using fuzzy labels. Fuzzy logic is used as means for introducing generalization in the state space and classifying input variables by using a sufficiently large but limited number of fuzzy sets. System states are determined during operation using fuzzy labels that result from a predefined set of rules describing each considered situation. Consequently, the implemented Fuzzy Inference System (FIS) (cf. [11]), assesses system conditions based on these classification results and decides upon which adaptations or counteractions to trigger by following its EEP, thus learning to optimize system performance given varying system states. In order to evaluate the severity of deviations from KPI targets, so-called *membership functions* [11] are introduced. The relative KPI error $e(t)$ of the considered KPI (e.g. CDR) is determined with respect to the observation time interval T as follows:

$$e(t) = \frac{KPI(t) - KPI^*}{KPI^*}, \quad (5)$$

where KPI^* denotes the performance target specified by the MNO policy (e.g. 1%). In principle, various types of membership functions (e.g. triangular, trapezoidal, or Gaussian shaped) can be used for classifying KPI performance. The sensitivity or "aggressiveness" of the classification scheme can be varied by tuning the membership function's shape, slope, or the amount of overlap with neighboring membership functions. In general, the decision on the different value ranges for each KPI state depends on observation time T and may also be influenced by MNO policy. Figure 2 illustrates an exemplary membership function that is used for assessing system performance with respect to KPIs. In the developed MRO scheme the respective

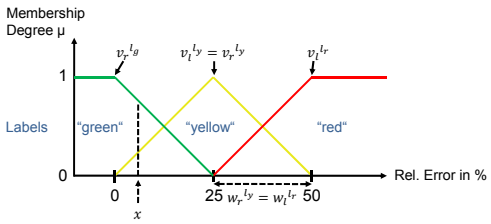

Fig. 2. Membership Functions for Assessing Relative KPI Error

KPI error classification results are used for triggering HO parameter adaptations instead of relying on fix timer and threshold values (cf. [2]). In order to perform inference, each rule is first quantified using fuzzy logic (FL) and then the applicability of each rule is determined. This is accomplished by comparing the premises of all rules to the FIS inputs and by determining which rules apply to the current situation. This *matching* process involves assigning the certainty that each rule applies, and typically yields a set of rules the FIS designer is more certain to apply to the current situation. For MRO, the designed FIS considers errors of several KPIs, such as CDR,

HFR, PHR, as well as the magnitude of the HYS parameter and employs a FL-based classification scheme for mapping HYS and KPI inputs s_j on fuzzy labels l_j^i. The rule set listed in table I summarizes the rule consequents and corresponding actions. A certain prioritization among these KPIs and their

TABLE I. FQL-BASED MRO RULE CONSEQUENTS AND ACTIONS

Rule	HYS	CDR	HFR	PHR	Actions
1	low	low	low	low	No action
2	low	low	low	$\neq$ low	{HYS ↑, HYS ↑ & TTT ↑, TTT ↑}
3	low	low	$\neq$ low	-	{HYS ↑ & TTT ↑, TTT ↑, HYS ↑}
4	low	$\neq$ low	-	-	{HYS ↓, HYS ↓ & TTT ↓, TTT ↓}
5	high	low	low	low	No action
6	high	low	low	$\neq$ low	{TTT ↑, HYS ↑ & TTT ↑, HYS ↑}
7	high	low	$\neq$ low	-	{TTT ↑, HYS ↑ & TTT ↑, HYS ↑}
8	high	$\neq$ low	-	-	{HYS ↓ & TTT ↓, HYS ↓, TTT ↓}

states is realized by assigning particular, MNO defined weights (e.g. $w_{CDR} = 4, w_{HFR} = 2, w_{PHR} = 1$). Further, the degree of counteracting indicated by the FIS conclusion depends on the deviation from the respective performance target. For example, if ping-pong HO performance degradation is classified 'medium' and the counteraction foresees to increase HYS parameter by one step, the step size is 0.5 dB. In contrast, if ping-pong HO performance degradation is assessed as 'high', the corresponding magnitude of HYS parameter adaptation would be 1 dB. In general, the number of rules $\mathcal{N}$ is determined by the number of input variables $\mathcal{N}_I$ (e.g. number of considered KPIs) and the number of fuzzy labels for each of these variables $\mathcal{N}_{L_j}$, i.e. $\mathcal{N} = \prod_{j=1}^{\mathcal{N}_I} \mathcal{N}_{L_j}$. Each of these rules has $\mathcal{N}_O$ corresponding conclusions $\left(o_m^i\right)_{m=1,\cdots,\mathcal{N}_O}^{i=1,\cdots,\mathcal{N}}$, that in the simplest case are crisp. Hence, the conclusion vector $\boldsymbol{o}_m$ can be used to approximate the output function y_m. Here, potential outputs are individual or joint adaptations of HYS and TTT settings in case of cell sector-specific adaptations, while in case of cell pair-specific optimization the CIO of the target eNB experiencing high KPI degradations is modified. Conclusions on which control actions to take are derived using the rules that have been identified to apply at the current time. A rule is *on at time* t, if its premise membership function $\mu_{premise} > 0$. The inference mechanism seeks to combine the recommendations of all the rules to come up with conclusions for HO parameter adaptations, that are characterized by fuzzy sets and represent the certainty that the HO parameters should take on various values. The certainty of a premise, composed of different linguistic terms and quantified by various membership functions, is determined using standard Boolean logic operators, such as *Minimum* or *Product*. The truth value of each rule i with respect to $\boldsymbol{s} \in \mathcal{S}$ can be computed as follows:

$$\alpha_i(\boldsymbol{s}) = \prod_{j=1}^{\mathcal{N}_I} \mu_{l_j^i}(s_j). \tag{6}$$

The defuzzification operation operates on the implied fuzzy sets produced by the inference mechanism that derived the activated rule truth values and combines their effects on the consequent membership functions $\mu(y_m)$ to provide the most certain FIS output. One popular method for combining the recommendations represented by the implied fuzzy sets from all activated rules is the so-called *center of gravity* (COG) defuzzification method:

$$y_m = \frac{\sum_i c_i^m \cdot \int \mu_i(y_m)}{\sum_i \int \mu_i(y_m)}, \tag{7}$$

where m denotes the corresponding FIS output, c_i^m denotes the center of the membership function of the consequent of rule i ($c \in \{-1, 0, 1\}$) and $\int \mu_i(y_m)$ denotes the area under the consequent membership function $\mu_i(y_m)$. The FQL scheme presented in this paper is based on the approach described in [12]. The *reinforcement* signal r received at time $t+1$ after applying the selected action(s) and leading the system to state $s' = s(t+1)$ consists of a weighted sum of KPI errors and can be written as follows:

$$r(t+1) = \sum_{m=1}^{N_{KPI}} w_m \cdot (KPI_m^* - KPI_m(t)), \tag{8}$$

where KPI_m^* denotes the respective KPI target values and $KPI_m(t)$ the observed and measured KPI at time t, respectively. $\{w_m\}$, $m \in \{1, \ldots, N_{KPI}\}$, represents the weights assigned to the respective KPIs specified by MNO policy. The reinforcement r will be positive if the KPIs are below the pre-defined KPI target values. With every KPI exceeding its performance target the numeric value of reinforcement r will be decreased and might become negative. Thus, the quality of the recently applied rule-action pair will be decreased, too. The rule quality values or briefly *Q-values* are calculated as follows:

$$Q(\boldsymbol{s}, \boldsymbol{a}) = \frac{\sum_{i=1}^{\mathcal{N}} \alpha_i(\boldsymbol{s}) \times q(\boldsymbol{s}_i, \boldsymbol{a}_i)}{\sum_{i=1}^{\mathcal{N}} \alpha_i(\boldsymbol{s})}, \tag{9}$$

where the function performing the mapping from input vector $\boldsymbol{s}$ onto rule consequents $\alpha_i(\boldsymbol{s})$ gives the truth value of rule i given input vector $\boldsymbol{s}$. The Q-values are updated according to:

$$\Delta Q = r(t+1) + \gamma \cdot V(\boldsymbol{s}(t+1)) - Q(\boldsymbol{s}(t), a), \tag{10}$$

where the concept of *eligibility* is applied that aims at reducing the so-called *temporal credit-assignment problem* [11]. For this purpose, so-called eligibility factors per rule-action pair $e(i,k)$ are introduced in order to memorize the applicability of each rule-action pair weighted according to their proximity to time step t. The eligibility factor of rule i and action k is defined as follows:

$$e(i,k) = \begin{cases} \lambda \cdot \gamma \cdot e(i,k) + \dfrac{\alpha_i(\boldsymbol{s})}{\sum_{i=1}^{\mathcal{N}} \alpha_i(\boldsymbol{s})} & k = k^+ \\ \lambda \cdot \gamma \cdot e(i,k) & \text{otherwise,} \end{cases} \tag{11}$$

where the *recency factor* λ, also called *eligibility rate*, is used to weight the formerly discounted eligibility value $\gamma \cdot e(i,k)$. Thereby, the updating equation becomes [12]:

$$\Delta q(i,k) = \kappa \cdot \Delta Q \cdot e(i,k). \tag{12}$$

where $\kappa \in (0,1)$ denotes the *learning rate*. Initially, all q-values $q(i,k)$ are zero. During learning process they are incrementally adjusted according to the experience attained through applying selected actions, while following a certain EEP. In the following, a limited EEP is assumed that is ε-*Greedy* in the sense that for each rule i the best action, referred to as k^* with $q(i,k^*)$, is selected with probability ε and random

action with a probability $(1-\varepsilon)$ [11]. More precisely, let k^+ be the selected action in rule i using an EEP and let k^* be the best action, i.e. $q(i,k^*) = \max_{j \in \mathcal{A}^i} q(i,j)$, the global Q-value of the inferred action $\boldsymbol{a}$ given input vector $\boldsymbol{s}$ is [12]:

$$Q(\boldsymbol{s},\boldsymbol{a}) = \frac{\sum_{i=1}^{\mathcal{N}} \alpha_i(\boldsymbol{s}) \times q(i,k^+)}{\sum_{i=1}^{\mathcal{N}} \alpha_i(\boldsymbol{s})}, \tag{13}$$

and the resulting state value is:

$$V(\boldsymbol{s}) = \frac{\sum_{i=1}^{\mathcal{N}} \alpha_i(\boldsymbol{s}) \times q(i,k^*)}{\sum_{i=1}^{\mathcal{N}} \alpha_i(\boldsymbol{s})}. \tag{14}$$

Learning stops, if ΔQ falls within the area of convergence $r_{opt} \pm \theta$, where $r_{opt} = \sum_{m=1}^{N_{KPI}} w_m \cdot KPI_m^*$, i.e. current KPIs remain zero.

TABLE II. FQL SCHEME

1.	Initialize Q-value LUT: $\forall i \in \{1,2,...,\mathcal{N}\}, \forall k \in \{1,2,...,\mathcal{N}_{\mathcal{A}^i}\}$ $q(i,k)=0$ and set time $t=0$. Repeat:
2.	Receive system state $\mathbf{s}(t) = (CDR(t),\ HFR(t),\ PHR(t))$.
3.	For each rule i select an action k with the EEP: $k = \arg\max_j q(i,j)$ with probability ε or $k =$ random $\{j,\ j=1,2,...,\mathcal{N}_{\mathcal{A}^i}\}$ with probability $1-\varepsilon$.
4.	Calculate inferred output using COG method.
5.	Determine its corresponding quality $Q(\boldsymbol{s},\boldsymbol{a})$.
6.	Execute action $a(\boldsymbol{s})$ at time t that leads the system to state $\boldsymbol{s}' = \boldsymbol{s}(t+1)$. Receive reinforcement $r(t+1)$.
7.	Calculate truth values $\alpha_i(\boldsymbol{s}(t+1))$ for $i \in \{1,2,\ldots,\mathcal{N}\}$.
8.	Determine the value of the new state: $V_t(\boldsymbol{s}(t+1)) = \sum_{i=1}^{\mathcal{N}} \alpha_i(\boldsymbol{s}(t+1)) \cdot \max_k q(i,k)$.
9.	Calculate the variation of the quality $Q(\boldsymbol{s},\boldsymbol{a})$ $\Delta Q = r(t+1) + \gamma \cdot V_t(\boldsymbol{s}(t+1)) - Q(\boldsymbol{s},\boldsymbol{a})$
10.	Update the elementary quality $q(i,k)$ of each rule i and action k $\Delta q(i,k) = \kappa \times \Delta Q \times e(i,k)$.
11.	Save the elementary quality $q(i,k)$ in the q-value LUT.
12.	If convergence is obtained, stop learning.
13.	$t = t+1$.

TABLE III. SIMULATION PARAMETERS

Parameter	Assumption
Carrier frequency	2 GHz
System bandwidth	10 MHz (50 PRBs)
Total transmit power	40 W
Control channel overhead	12%
Shadowing	log-normal Standard deviation: 8 dB Decorrelation distance: 50 m
Fast fading	2-tap Rayleigh fading channel
Noise power	-174 dBm/Hz + $10 \cdot \log_{10}(B) + 7$ dB
Background users per cell sector	5
High-speed users on highway lanes	780 at 120 km/h
Traffic model	CBR traffic, full buffer, user data rate = 192 kbps
HO preparation time	250 ms
RLF timeout (T310 [10])	1s
Ping-pong HO time	5 s
KPI update interval	1 s
KPI targets $(\{CDR^*, HFR^*, PHR^*\})$	$\{1.25\%, 2.5\%, 5\%\}$
KPI weights $(\{w_{CDR}, w_{HFR}, w_{PHR}\})$	$\{4, 2, 1\}$

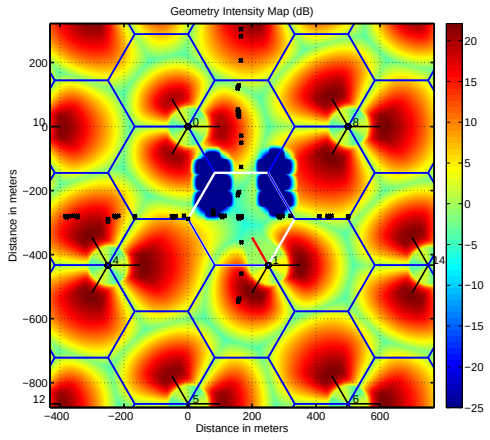

Fig. 3. Geometry Intensity Map and Cell Layout of Reference Scenario

IV. PERFORMANCE EVALUATION

A. Evaluation Methodology

For studying the performance of the developed FQL-based MRO scheme, a LTE system level radio network simulation tool is used, where following 3GPP's guidelines [13] a typical macro cell deployment with 3 sectors and site-to-site distances of 500 m are assumed. Table III summarizes simulation parameters.

Further, 5 slowly moving so-called *background* users are located in each cell sector in order to generate a basic traffic load, besides the 780 high speed users that are foreseen to continuously move at 120 km/h on several highway lanes (3 per direction), illustrated in Fig. 3 as black crosses. Moreover, simplified *coverage holes* (in dark blue) are intentionally inserted in the depicted cell area overlaying the shadowing map in order to create possibilities for KPI degradations, thus triggering SON functionalities. In addition, the cell sector highlighted in white is switched off after 5% of simulation time (antenna steering direction in red) in order to evaluate impact on self-optimization mechanisms operating at neighboring cell sectors. A challenging, and more realistic user distribution consisting of convoys of e.g. a bus with 30 people followed by a number of cars (e.g. 3 cars with 2 passengers each) is modeled. Users that are about to leave evaluation area will enter again driving on the opposite lane. For modeling user service request, a simple full-buffer traffic emulation is used. The FQL-based MRO scheme operates independently at each cell sector, where KPI update interval is set to 1 s. The adaptation steps of HYS and TTT values follow 3GPP specifications. Hysteresis values range from 0 to 10 dB, where adaptations occur in steps of ± 0.5 dB. However due to the simulation tool's time basis being 10 ms, only multiples of 10 ms can be chosen as TTT value, limiting the TTT value set to $\{0, 0.04, 0.08, 0.1, 0.16, 0.32, 0.48, 0.64, 1.280, 2.560, 5.120\}$ s. The implemented FIS rule set and the limited EEP are listed in Tab. I, where the presumable best action given system state $\boldsymbol{s}$ corresponds to the first action column entry that is selected with a probability of $\varepsilon = 0.8$, while otherwise the two alternative actions are equally probable. The discount factor is set to $\gamma = 0.95$, the learning to $\kappa = 0.01$, the eligibility rate to $\lambda = 0.9$, and the convergence limit to $\theta = 0.001$, respectively. Performance targets of the different KPIs (CDR, HFR, PHR) are set to $\{1.25\%, 2.5\%, 5\%\}$. Furthermore, overall optimization performance is evaluated using *overall performance indicator* (OPI) that is calculated

as follows:

$$OPI = \sum_{m=1}^{N_{KPI}} w_m \cdot \frac{KPI_m^* - KPI_m(t)}{KPI_m^*}. \qquad (15)$$

The optimization task to be performed by each FQL entity is to maximize the value of OPI, where here $OPI_{opt} = \sum_{m=1}^{N_{KPI}} w_m \cdot KPI_m^* = 7$ given KPI target and weight values.

B. Simulation Results

For benchmarking the performance of the designed FQL-based MRO scheme, the trend-based MRO algorithm presented in [2], and a scheme that assigns TTT values based on velocity estimates (*VC_TTT*), which may be obtained via Doppler shift measurements, was implemented, too. In the first case, two different configurations were considered, which are referred to as *TBHOA* and *TBHOA2*. For *TBHOA*, bad/good performance timer values of 1/2s and for *TBHOA2* 5/10s are used, respectively. In case of *VC_TTT*, fix TTT values of 320 ms are set for high speed users, whereas in the reference case *REF* a fix TTT setting of 160 ms is considered. Fig. 4 depicts the connection dropping ratios CDR for each neighboring cell sector of the switched off cell sector, e.g. due to hardware malfunction, highlighted in white in Fig. 3.

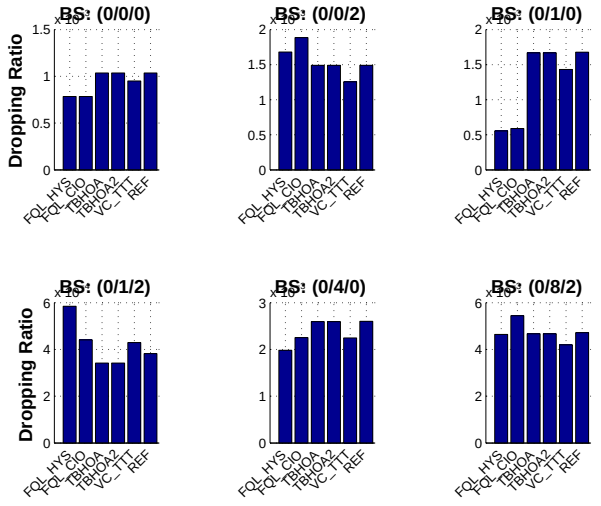

Fig. 4. Connection Dropping Ratios

Except the cell sectors northwest $(0/0/2)$ and south $(0/1/2)$ of the switched off sector, the FQL-based optimization approaches are able to enhance CDR performance, in particular with respect to sector $(0/1/0)$. The trend-based HO optimization approaches only outperform other schemes in case of cell sector $(0/1/2)$. Although not primarily triggered due to CDR or HFR degradations, most significant performance gains through exploiting learned HO parameter adaptations are achieved with respect to HFR, as illustrated in Fig. 5.

Only for cell sector $(0/8/2)$, the FQL optimization scheme tuning CIO parameters of neighboring cell sectors is not able to achieve same performance as the hysteresis-based approach, but even yields the second worst performance, which is still below its KPI target of 2.5%. Applying *TBHOA* and *TBHOA2* in cell sector $(0/1/0)$ results in increased HO failures, too. In terms of PHR performance, optimization is triggered due

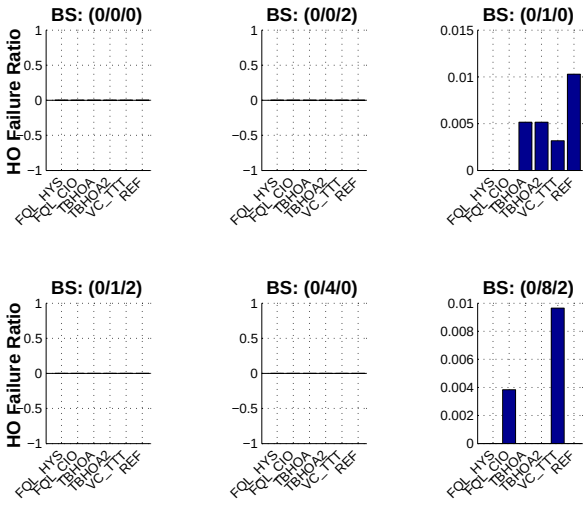

Fig. 5. Handover Failure Ratios

to KPI target violation in all sectors except $(0/0/2)$ and $(0/4/0)$, where only in sector $(0/4/0)$ *VC_TTT* is able to improve PHR performance. In all other cases, the FQL-based approaches outperform other schemes, where the CIO-based scheme performs best among all, as depicted in Fig. 6.

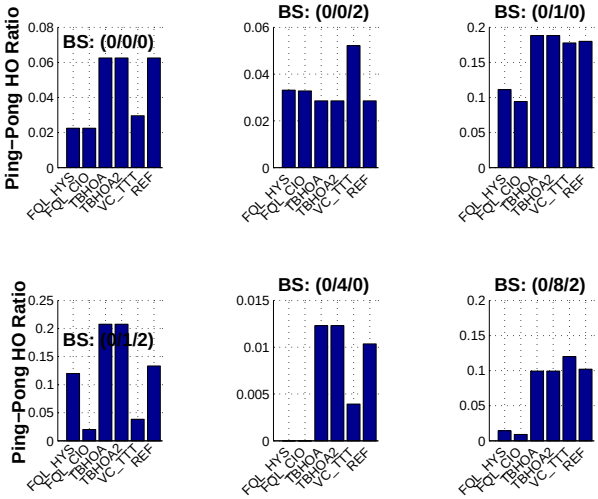

Fig. 6. Ping-Pong Handover Ratios

Evaluating overall performance of all optimization schemes using OPI, as defined in Eq. 15 and illustrated in Fig. 7, the FQL-based approaches outperform other approaches except in case of sector $(0/0/2)$, where neither FQL-based nor *TBHOA* or *TBHOA2* were triggered. Further, the FQL-based approach, where hysteresis values are tuned in a cell-specific manner achieves best performance, since the KPI CDR is assigned highest weight, e.g. due to MNO policy.

Fig. 8 depicts the sum of too early handovers, where significant and almost stable reductions are achieved by *FQL_HYS* over several training episodes, while the performance results of *FQL_CIO* are oscillating around those of *VC_TTT*. Too late or wrong cell handovers were not observed in the considered scenario. User experience and QoS are assessed based on considered traffic model (cf. Tab. III), where the overall number of

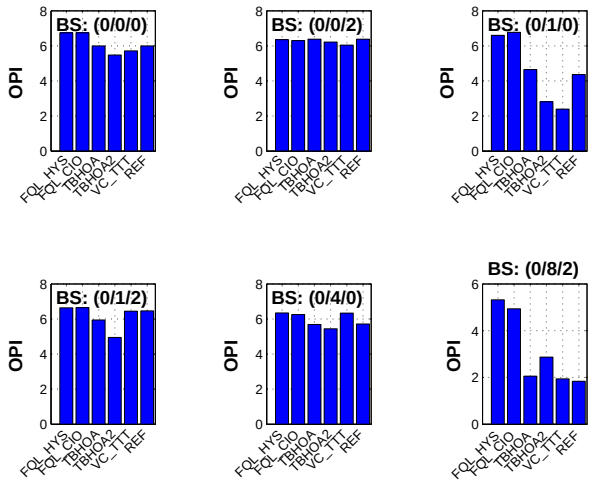

Fig. 7. Overall Optimization Performance Indicator

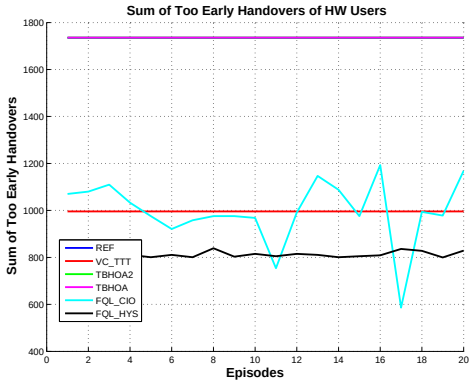

Fig. 8. Sum of Too Early Handovers

users was evaluated that were provided the requested service bandwidth. The overall number of satisfied users is depicted in Fig. 9. Except cell sector $(0/0/2)$, where neither FQL-

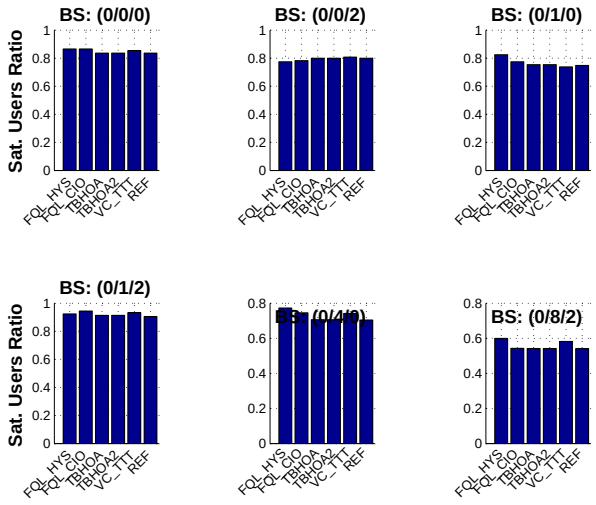

Fig. 9. Sum of Satisfied Users

based nor trend-based approaches were triggered, FQL-based schemes clearly outperform trend-based approaches, whereas gains with respect to *VC_TTT* are lower.

V. Conclusion

The presented FQL-based MRO scheme enables self-tuning of HO parameters using a limited rule set that reflects the situations the system may be exposed to. It learns optimum parameter adaptations using an EEP. Further, this scheme is able to simultaneously account for several KPI performance targets and by increasing its knowledge base self-optimize its operation. Benchmark results show that the FQL-based scheme clearly outperforms other considered schemes in particular with respect to HO failures and ping-pong handovers, where optimum performance was achieved. However, performance enhancements could not be achieved with respect to all cell sectors, since defined KPI targets were set too conservative. Further, modifying assigned KPI weights used in the FQL-based schemes allows for prioritizing different KPIs depending on eNB location and operation conditions. Learning to set and adapt performance targets and KPI weights according to locally observed conditions and KPI evolutions, is another topic for future work.

Acknowledgment

Part of this work has been performed in the framework of FP7 project ICT-317669 METIS, which is partly funded by the European Union. The authors alone are responsible for the content of the paper.

References

[1] Nokia Siemens Networks, “2020: Beyond 4g radio evolution for the gigabit experience,” Aug. 2011, white Paper.

[2] T. Jansen, I. Balan, J. Turk, I. Moerman, and T. Kürner, “Handover Parameter Optimization in LTE Self-Organizing Networks,” in *72nd IEEE Vehicular Technology Conference (VTC-Fall)*, Sept. 2010, pp. 1–5.

[3] Next Generation Mobile Networks (NGMN) Alliance, “Use Cases related to Self Organising Network, Overall Description,” May 2007.

[4] ——, “Recommendation on SON and O&M Requirements,” Dec. 2008.

[5] J. Alonso-Rubio, “Self-optimization for handover oscillation control in LTE,” in *IEEE Network Operations and Management Symposium (NOMS)*, April 2010, pp. 950–953.

[6] L. Ewe and H. Bakker, “Base station distributed handover optimization in LTE self-organizing networks,” in *22nd International IEEE Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept. 2011, pp. 243–247.

[7] K. Kitagawa, T. Komine, T. Yamamoto, and S. Konishi, “A Handover Optimization Algorithm with Mobility Robustness for LTE Systems,” in *22nd IEEE International Symposium onPersonal Indoor and Mobile Radio Communications (PIMRC)*, Sept. 2011, pp. 1647–1651.

[8] I. Viering, B. Wegmann, A. Lobinger, A. Awada, and H. Martikainen, “Mobility robustness optimization beyond Doppler effect and WSS assumption,” in *8th International Symposium on Wireless Communication Systems (ISWCS)*, Nov. 2011, pp. 186–191.

[9] Z. Ren, P. Fertl, Q. Liao, F. Penna, and S. Stanczak, “Street-specific handover optimization for vehicular terminals in future cellular networks,” in *77th IEEE Vehicular Technology Conference (VTC-Spring)*, 2013.

[10] 3GPP TSG RAN, “E-UTRA Radio Resource Control (RRC) (Release 11),” Protocol Specification 36.331 V11.1.0, September 2012.

[11] L. Jouffe, “Fuzzy Inference System Learning by Reinforcement Methods,” *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 28, no. 3, pp. 338–355, Aug. 1998.

[12] P. Glorennec and L. Jouffe, “Fuzzy Q-learning,” in *6th IEEE International Conference on Fuzzy Systems*, vol. 2, July 1997, pp. 659–662.

[13] 3GPP TSG RAN, “Further advancements for E-UTRA physical layer aspects (release 9),” Technical Report 36.814 V9.0.0, March 2010, available online at http://www.3gpp.org/ftp/specs/html-INFO/36814.htm.