Optimal Load Management with inclusion of Electric Vehicles and Distributed Energy Resources

F. Milano, *Senior Member, IEEE*, O. Hersent

Abstract— This paper presents an optimal load management (OLM) strategy for distribution systems with high penetration of electric vehicle chargers and distributed energy resources. The paper describes two formulations of the OLM and practical implementation issues aimed to improve OLM accuracy and to reduce its computational burden. The proposed OLM formulations are specifically designed to provide a good compromise between computational burden and result accuracy. A radial system with three-phase unbalanced section is used for illustrating the proposed OLM technique.

Index Terms— Load Management, electric vehicle (EV), distributed energy resources (DER), real-time control.

I. INTRODUCTION

A. Motivation

IN recent years, the development and diffusion of electric vehicles as well as of distributed energy resources has grown exponentially. The high penetration of these two technologies forces to rethink how distribution systems are handled. In particular, an open challenge is how to properly feed electric vehicle chargers while maintaining security and quality standards. This paper presents an optimal load management (OLM) strategy for distribution networks with high penetration of electric vehicle (EV) chargers and distributed energy resources (DERs). The design of the OLM has two primary goals: accuracy and efficiency. Accuracy is needed to properly decide whether a load request can be admitted without leading to the violation of technical limits, such as under-voltages and/or distribution line current limits. The OLM must be able to properly and precisely compute and recognize limit violations. Efficiency is needed to ensure that load and DER requests are processed in real-time and that loads are not queued unnecessarily.

B. Literature review

The literature on load management is vast and cover various aspects from electricity pricing to on-line applications. For example, [1] and [2] propose real-time load management techniques that are able to minimize the total cost of energy generation as well as network losses. In [3], the authors

propose an optimization model to adjust the hourly load level of a given consumer in response to hourly electricity prices. In [4], an OPF problem based on smart metering is defined to minimize load shedding while maintaining system security. A load response and energy management based on agents is proposed in [5]. Electric vehicles pose new challenges to the operation of distribution systems. In [6], the author proposes a charging method for plug-in hybrid electric vehicles with a user demand regulation based on pricing information. In [7], a unidirectional vehicle-to-grid optimal charging strategy is proposed to maximize aggregated profits. References [8]–[11] illustrate the impact of electric vehicles on distribution systems with particular regard to their effect on system security. Finally, the inclusion of distributed energy sources and their impact on system security is another open question. Recent works on this topic are, for examples, [12] and [13]. In particular, [12] presents a dynamic modeling and control strategy for a sustainable micro-grid powered by wind and solar energy, while [13] presents an impact analysis of distributed energy resources integration on distribution systems, focusing mainly on reliability aspects. In [14] the usage of electric vehicle as distributed energy storage systems for system with high wind power penetration is discussed.

In this paper, the load management is formulated as an on-line software tool that recollects load and distributed energy resources (DER) requests from the system and decides whether a given request is acceptable for the system or not. If it is acceptable, the OLM returns to the system proper signals so that the load/DER is admitted, otherwise, the load/DER remains in a queue, waiting for being admitted. The goal of the OLM is to maximize the number of admitted power requests (integer variables) subject to system power flow (nonlinear) constraints and technical limits. Thus, the natural way to formulate the OLM is as a mixed integer nonlinear programming (MINLP) optimization problem. However, there is currently no well-assessed, fast and robust algorithm able to solve MINLP problems. Existing solvers have to be properly tuned to provide a solution in a reasonable time and never guarantee the global optimum (see for example the discussion in [15]). The interested reader can find some of these techniques in [16]–[18]. Unfortunately, the computational burden of MINLP solvers precludes their usage for on-line applications. Hence, the two proposed OLM formulations are based on a quadratic programming optimization problem and heuristic rules characterized by a reduced computational burden. The proposed optimization problems basically implements a *fair admission rule* (FAR) that is aimed to reschedule, in case of congestion, existing electric vehicle power requests. An in-depth discussion on theoretical and practical aspects of the implementation
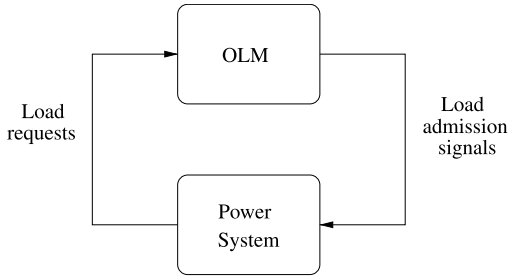
Fig. 1. Qualitative illustration of the OLM.

of the FAR is given in the paper.

C. Contributions

In summary, the contributions of this paper are twofold:

1) The formulation of two optimal power flow problems to properly handle load and DER request queues while respecting technical limits of the network and avoiding congestion.
2) The proposal of a fair admission rule able to satisfy as many load requests as possible while avoiding greedy requests.

Both the optimal power flow problems and the fair admission rule show a light computational burden and are thus suitable for on-line applications.

D. Paper Organization

The paper is organized as follows. Section II describes the hypotheses, the structure and the organization of the proposed optimal load management. Section III presents an in-depth discussion of the proposed OLM routine based on a 50-bus radial network. Section IV draws conclusions and indicates future work directions.

II. Proposed Optimal Load Management

The proposed OLM is an on-line software application that decides whether a certain load request is acceptable or not. If it is acceptable, the OLM sends the proper signal to the system (e.g., to an EV charger) so that the load is fed, otherwise, the load request remains in a queue, waiting for being admitted. According to the previous description, the OLM can be pictorially illustrated through the scheme shown in Fig. 1.

In this context, the term *acceptable* means that the load request, if admitted, does not lead to technical limit violations, such as over or under voltages, and transmission line or transformer overloads (i.e., over-currents). The power flow analysis is the adequate tool to define whether a certain load power consumption lead to limit violations or not. Observe that, since the power flow problem requires the solution of a set nonlinear equations, it is not possible to take advantage of the *superposition principle*. In other words, the effect of adding a new load power consumption at a certain bus cannot be known without solving again the power flow problem for the whole network. The non-linearity of the power flow problem is the major issue of the OLM.

According to Fig. 1, the OLM interacts with the power system by exchanging information. Let assume that such information is not continuous, but updated at a certain sample rate, e.g., 1 s or 5 s. During this time interval, the OLM must process the input signals from the network (i.e., the load requests), and decide which loads can be admitted and which ones have to wait. Then, at the following update, the OLM has to send the load admission signals to the system and receive the new load requests.

The functioning of the communication system is out of the scope of this paper. The hypothesis is that the communication system works properly and that the assumed sample interval is sufficient to send and receive all signals to and from the OLM, respectively.

Let assume that at a given snapshot t, m power requests reach the OLM. These requests can be either new or old ones that were queued in a previous step. The OLM has also to take into account that p active devices, i.e., loads or DERs that are currently connected to the grid, are discontinued at a given snapshot t_i, i.e., when the supply period is over. Let m_a the number of loads that can be accepted at each time t, with $m_a \leq m$. The objective of the OLM is to maximize the number m_a of load requests that are satisfied at a certain snapshot t.

The proposed OLM algorithm flow is described below.

1) Assume a feasible initial operating point, i.e., a point for which no technical limit is binding. This assumption is reasonable, as if one assumes that the OLM is always in service, no limit load request leading to a limit violation can be admitted.
2) Collect new load requests and add them to the pending load request queue. Moreover, discard previously admitted loads that have completed their connection period and pending load requests that have waited for a given period.
3) Sort load requests based on load priority levels (see definition below). In case of two load requests have same priority level, they are sorted based on the time waited to be served. Finally, in case also the waited time is the same, priority is given to load requests with higher energy demand.
4) Run one power flow per each load request in the order defined by the sorting process of Step 3. If for a certain load request, the power flow converges and no limit is violated, then the load is admitted. Otherwise the routine applies a *Fair Admission Rule* (FAR) that attempts to reschedule load power consumption and connection periods in order to accept the new load request. If all previous attempts fail, the load request is put in the queue of loads waiting for being supplied. This step is illustrated in Figure 2.
5) Go back to step 2.

The *priority level* is a concept that we introduce to discriminate among load requests. The priority level is basically an index or a flag that is assigned to each load request. In our vision, priority levels are based either on the price of charge
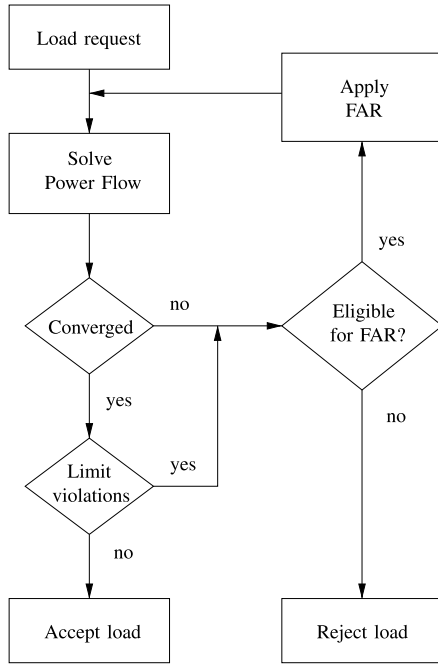
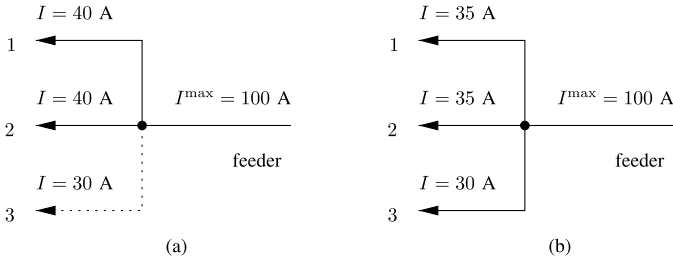Fig. 2. Flowchart of the admission/rejection process of load requests.

Fig. 3. Illustration of the proposed fair admission rule. Assume that loads 1 and 2 are preexisting admitted loads. (a) Load 3 cannot be admitted not to violate the maximum current limit. (b) Load 3 is admitted by reducing load 1 and 2 currents.

or on social considerations (e.g., a vehicle whose owner is a doctor in service receives the highest priority). The priority level is hence a signal associated to the load request. Of course, other sorting criteria can be implemented or added to the ones indicated in point 3 above (e.g., the charging modes implemented by Nissan). However it is worth observing that the general conclusions that can be drawn for the proposed OLM procedure does not rely on the specific criterion used to sort load requests.

As discussed above, the FAR is a procedure that attempts to reschedule loads in order to allow all load requests to be satisfied. The main idea is that all electric vehicles can be supplied, even if at a lower current level than requested and, thus, for a longer period than initially scheduled. The FAR is illustrated in Fig. 3.

The problem to be solved by the FAR can be posed as follows. Assume that a new power request p_k raised by a certain load k causes the violation of some technical limits (either bus voltage or transmission line current limits). We are interested in finding a "fair" rescheduling of all admitted loads

$i = 1, \ldots, n$ (including the new load request k) such that the power flow solution does not show any congestion or limit violation.

Observe that the FAR introduces a loop in the load request process. The loop breaks only when the load rescheduling leads to a power flow solution without limit violations and congestion. The existence of such condition is always guaranteed since the load profile prior to the new load request admission is feasible. Observe also that DER offers are treated in a similar way as load requests.

A. Fair Admission Rule (FAR)

This section describes in detail theory, implementation and practical aspects of the proposed FAR.

The conditions to take into account to properly reduce loads are the following:

1) Loads with lower priority levels have to be rescheduled before loads with higher priority levels. Only in case, the rescheduling cannot be completed using low priority loads, high priority loads can be rescheduled.
2) Loads that most "participate" to the congestion have to be penalized more than loads that are not responsible of the congestion. "Participation" is evaluated in terms of sensibility coefficients, i.e., the derivative of the binding limit with respect to load power consumption.
3) Loads cannot be reduced indefinitely. The threshold of the maximum admissible reduction is fixed by the OLM operator.
4) Finally, the goal of the FAR is to reduce the minimum amount of load powers such that the new power request can be satisfied.

The rules and conditions described above can be conveniently formulated as a quadratic programming optimization problem as follows:

$$\underset{(\Delta \boldsymbol{p})}{\text{minimize}} \quad \frac{1}{2} \sum_{i=1}^{n} w_i \Delta p_i^2 \tag{1}$$

$$\text{subject to} \quad \Delta p_{\text{tot}} = \sum_{i=1}^{n} \Delta p_i \qquad\qquad : \lambda$$

$$0 \leq \Delta p_i \leq \Delta p_i^{\text{max}} \quad : \pi_i^{\text{max}}, \quad i = 1, \ldots, n$$

where:

Δp_i : power reduction of load i in kW.
Δp_{tot} : total desired power reduction in kW.
w_i : weighting factor for load i.
Δp_i^{max}: maximum allowed power reduction of load i in kW.
λ : Lagrangian multiplier of the equality constraint.
π_i^{max} : Lagrangian multiplier of inequality constraints.

Problem (1) can be solved using either quadratic programming (QP) solvers (e.g., CPLEX, MOSEK, etc.) or cone programming (CP) solvers (e.g., CVXOPT) being the main computational burden due to the inequality constraints.

Since one of the main technical constraints of the OLM is to be suitable for on-line applications, we propose an alternative approximated and iterative formulation that can replace (1).

The simplified QP problem is the following:

$$\underset{(\boldsymbol{\Delta p})}{\text{minimize}} \quad \frac{1}{2}\sum_{i=1}^{n} w_i \Delta p_i^2 \tag{2}$$

$$\text{subject to} \quad \Delta p_{\text{tot}} = \sum_{i=1}^{n} \Delta p_i \quad : \lambda$$

where inequalities have been removed. The main advantage of (2) with respect to (1) is that the former has an explicit solution, which can be obtained imposing the Karush-Kuhn-Tucker (KKT) optimality conditions to the Lagrangian function of (2). The Lagrangian function is:

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{n} w_i \Delta p_i^2 - \lambda\left(\sum_{i=1}^{n}\Delta p_i - \Delta p_{\text{tot}}\right) \tag{3}$$

Then the KKT conditions are:

$$0 = \frac{\partial \mathcal{L}}{\partial \Delta p_i} = w_i \Delta p_i - \lambda, \quad i = 1,\ldots,n \tag{4}$$

$$0 = \frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{n}\Delta p_i - \Delta p_{\text{tot}} \tag{5}$$

Hence, the reduction in admitted power of load i is:

$$\Delta p_i = \frac{\lambda}{w_i} \tag{6}$$

where the Lagrangian multiplier λ is:

$$\lambda = \frac{\Delta p_{\text{tot}}}{\sum_{i=1}^{n} 1/w_i} \tag{7}$$

Observe that the explicit solution of (2) implicitly imposes that $\Delta p_i \geq 0$ but, of course, it cannot guarantee that the conditions $\Delta p_i \leq \Delta p_i^{\max}$ are satisfied. Hence the FAR requires an iterative solution of problem (2), as follows:

1) At each new snapshot, the weighting factors w_i and the total power to be shed Δp_{tot} are computed.
2) Then problem (2) is solved, thus obtaining load power reductions Δp_i for each load i.
3) If all $\Delta p_i \leq \Delta p_i^{\max}$, the FAR rule terminates. Otherwise, we set $\Delta p_i = \Delta p_i^{\max}$ to all loads for which the condition $\Delta p_i \leq \Delta p_i^{\max}$ is not satisfied and the associated weighting factor is set to $w_i = 0$ (so that those loads do not participate to the following power rescheduling).
4) The routine continues at point 2 and iterates until either the full amount Δp_{tot} is shed or all weighting factors are $w_i = 0$. In the latter case, all load powers are reset to their initial values and the problem is considered unfeasible. This mean that the current power request k is rejected and queued.

The algorithm described above is illustrated in the flowchart of Fig. 4. Observe that the solution obtained by the iterative technique depicted in Fig. 4 is quasi-optimal as only problem (1) provides optimal solutions. However, since in general $\Delta p_{\text{tot}} \ll \sum_{i=0}^{n} p_i$, in most cases constraints $\Delta p_i \leq \Delta p_i^{\max}$ are not binding and thus the solutions of (1) and (2) coincide.

Both (1) and (2) rely on some parameters that have to be tuned by the operator and thus introduce a certain degree of
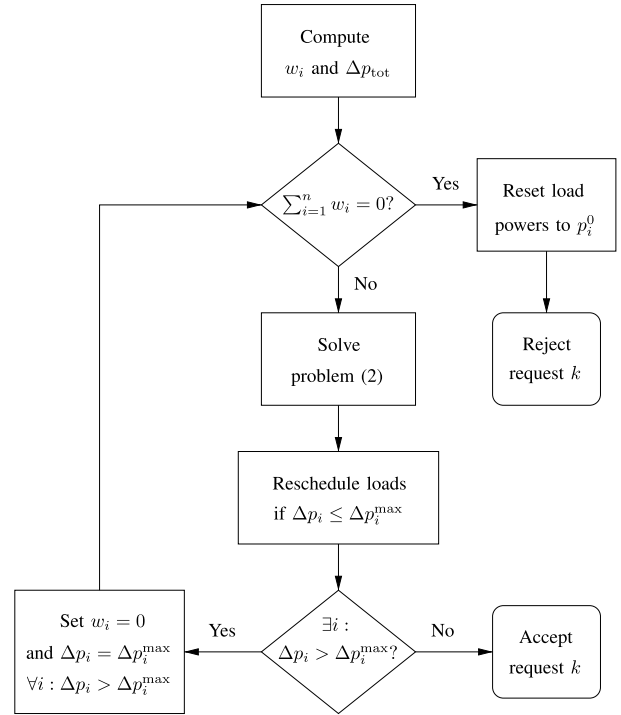

Fig. 4. Implementation of the FAR consisting in repeating iterative solution of the quadratic programming problem (2).

arbitrariness into the power rescheduling. Parameters to be carefully chosen are the weighting factors w_i for each load i, the maximum allowable power $\Delta p_i^{\max}$ that can be shed for each admitted load, and the total desired load shedding Δp_{tot}. A discussion on how to chose each of these parameters is given below.

Weighting factors w_i:
Weighting factors have to be chosen so that the lower the priority level of the load i and the bigger the sensitivity of load i to the binding limit, i.e., the higher the sensitivity of the binding limit with respect to the power p_i, and the higher the value of p_i, the smaller the weighting factor w_i. Based on the considerations above, we propose the following expression for w_i:

$$w_i = \left(\frac{1}{\sigma_i \cdot (\ell^{\max} - \ell_i) \cdot (p_i^0 / \sum_{j=1}^{n} p_j^0)}\right)^2 \tag{8}$$

where:
σ_i is the sensitivity of the binding limit with respect to load power p_i. Hence, one has:

$$\sigma_i = \frac{\partial i_h}{\partial p_i} \tag{9}$$

or

$$\sigma_i = \frac{\partial v_j}{\partial p_i} \tag{10}$$

which define the sensitivity of the current in transmission line with index h and of the voltage magnitude at bus j, respectively, with respect to power p_i.
$\ell^{\max}$ is the maximum priority level that can be assigned to a load.

ℓ_i is the priority level of load i.

p_i^0 is the power currently admitted for load i. In case $i = k$, p_i^0 is the current power request of load k.

Substituting (8) in (6) leads to:

$$\Delta p_i = \left(\sigma_i \cdot (\ell^{\max} - \ell_i) \cdot \frac{p_i^0}{\sum_{j=1}^n p_j^0} \right)^2 \cdot \lambda \qquad (11)$$

which duly satisfies the properties of the expected solution.

The following are relevant remarks on the proposed definition of the weighting factors w_i.

- If $p_i^0 = 0$, i.e., the load is currently not admitted, then $\Delta p_i = 0$, as expected.
- If $\sigma_i = 0$, then $\Delta p_i = 0$, as expected, as the variation of p_i does not have any effect on the binding limit.
- If $\ell_i = \ell^{\max}$, the load i has maximum priority and cannot be modified as $\Delta p_i = 0$. If one wants to impose that all load can be rescheduled, then it suffices that $\ell_i < \ell^{\max}$.
- The higher p_i, the bigger Δp_i. This is also a desired effect, as it is desirable that "greedy" loads are penalized more than non-greedy ones.
- The higher σ_i or the lower ℓ_i, the bigger Δp_i, as expected.
- If $w_1 = w_2 = \cdots = w_n$, then $\Delta p_i = \Delta p_{\text{tot}}/n$, $\forall i = 1, \ldots, n$, as expected.

Observe that the square for the parameters used for defining w_i is used to "amplify" the effect of each parameter. The rationale behind the square function is that the relative difference between the weighting factors of loads with different features has to be high. For example, assuming same sensitivity σ_i and same admitted power p_i^0, we want that if $\ell_i < \ell_j$, then $w_i \ll w_j$, and hence, load i is more likely to be rescheduled than load j.

Since the proposed technique is applied exclusively to radial networks, there is no need to deal with the signs of sensitivities σ_i. In radial distribution systems, current flows can be only in one direction, i.e., towards the loads. Hence, if there is a congestion on a certain branch, the sign of all σ_i is always positive. The same rationale applies to the sensitivities of DERs. The only condition that can impose to reduce DERs power injections is the case for which DER power productions are leading to a over-current in the transmission line. Hence, if DERs sensitivities are to be used, then they are certainly negative. Moreover, since in (8) weighting factors are computed through the square function, the effect of σ_i is always the expected one.

Finally, observe that, in (8), the absolute values of the factors σ_i and $(\ell^{\max} - \ell_i)$ does not affect the solution of the optimization problems (1) and (2). However, the relative values of weighting factors do matter. In fact, two loads having same sensitivities but different priority levels are curtailed in such a way that the higher the priority level ℓ_i, the lower Δp_i. In the same vein, two loads with same priority levels but with different sensitivities will be curtailed in such a way that the higher the sensitivity σ_i, the lower Δp_i.

Maximum allowable power $\Delta p_i^{\max}$:
A simple but fair criterion is to define $\Delta p_i^{\max}$ a limit proportional to the currently admitted power:

$$\Delta p_i^{\max} = \gamma_i p_i^0 \qquad (12)$$

where $0 < \gamma_i < 1$ must hold. However, it appears reasonable to assume $0.1 \leq \gamma_i \leq 0.25$.

Total desired load shedding Δp_{tot} :
Similarly to the parameters $\Delta p_i^{\max}$, the total amount of load to be shed is an arbitrary quantity. It can be defined based on the amount of the power request p_k that is causing the congestion:

$$\Delta p_{\text{tot}} = \gamma_{\text{tot}} p_k \qquad (13)$$

where $[0.1, 0.25]$ appears to be a reasonable range for γ_{tot}.

Although the coefficients γ_i are not related to γ_{tot}, all simulations discussed in the case study assume, without lack of generality, $\gamma_i = \gamma_{\text{tot}} = \gamma$. The value of either γ_i or γ_{tot} is based on heuristic considerations. The effect of different values of γ is illustrated in Section III.

B. Remarks on Customer Convenience and Priority Levels

In this subsection we provide some remarks on priority levels and show how these contribute to customer convenience.

- The priority level is selected by the customer and allows discriminating among customers that accept to be curtailed and thus accept a longer charging time, and customers that are not willing to wait longer than indicated at the connection time. If we assume that high priority levels cost more than lower ones, the customer can choose whether to pay more and wait less or to pay less and wait more. This approach incentivizes to better distribute the charging of electric vehicles during the day and, thus, to reduce congestion.
- Load power variations determined by the OLM are bounded to few percents of the original load request. This means that load power reductions lead to a marginal increase in the total charging time. In any case, as indicated in the previous point, users with strict charging time requirements can reduce or eliminate shedding by requiring a higher priority charging.
- In case of network congestion, the OLM operates so that the maximum number of customers can be served. While some customers (and only those that have allowed that) will have to wait for a slightly longer time than anticipated, the number of customers that have to wait to be served is minimized. Hence, the average convenience of all customers is increased by the proposed OLM technique.
- Finally, the proposed FAR allows connecting to the network, even in case of peak load conditions, special customers such as doctors' vehicles or any other customer that requires a special treatment.

C. Remarks on Sensitivities

The computation of sensitivities is the most demanding part of the FAR. There is no closed formula to compute the sensitivity of a line current or a bus voltage magnitude with respect to power consumption at buses. So the only way to compute sensitivities is to use a numerical approximation.

Since the solution for the current operating point is known, each sensitivity requires the solution of an additional power flow. For each load variation and hence for each power flow solution, one can obtain the transmission line current sensitivities, i.e., $\partial i_h/\partial p_i$, and bus voltage sensitivities, i.e., $\partial v_k/\partial p_i$, with respect to load powers. However, all these sensitivities are hardly needed all together as, typically, each new power request leads to only a few limit violations (typically only one).

Practical considerations allow simplifications in time and in location, as discussed below.

1) Time reduction: One of the main assumption of the whole OLM algorithm is that each new load request is relatively small with respect to the current total power consumption as well as with respect to the load consumption of the network section where the new load request is located. So, it is reasonable to assume that, even though power flow equations are nonlinear, each new load request does not vary "too much" the power flow solution and, hence sensitivities.

Thus, it appears reasonable to update sensitivities with a lower rate than the OLM routine time step. The appropriate frequency at which refreshing sensitivities is not known *a priori* and has to be adjusted based on the knowledge of the system.

However, a good index of the need to refresh sensitivities can be the number of "events" (e.g., new load admissions) that have occurred in the network since the last update. Clearly, one has to expect that the less the sensitivities are updated, the lower the accuracy of the FAR. Thus, there is a threshold between accuracy and simulation speed up.

2) Location reduction: In several mathematical applications, it has to be expected that solving a single huge problem with several variables is computationally more expensive than solving several smaller problems with few variables. This is certainly true if the computational complexity of the problem to be solved is polynomial. Location reduction attempts to decompose the QP problem solved for the full network into a set a smaller problems for network sections.

The common sense indicates that power variations of loads that are located far away (in an electrical sense) from a certain transmission line or a certain bus bar, are very likely going to show small or negligible sensitivities with respect to the current in the transmission line and to the voltage magnitude of the bus.

This intuitive notion of "electrical distance" can be easily exploited in radial distribution networks as different branches are likely quite decoupled from each other.

The concept of *tree* is introduced in the OLM routine to reduce the number of sensitivities to be computed for each limit violation, as follows. Network branches are grouped into topological trees, which are defined based only on topological
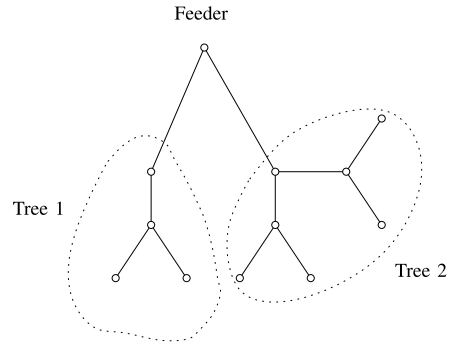

Fig. 5. Pictorial representation of network trees. The assumption is that load variations in Tree 1 do not affect currents and voltages of transmission lines and voltages pertaining to Tree 1. Observe that this hypothesis is fully satisfied only if the voltage at the feeder is constant.

considerations. Observe that the independence of tree sensitivities is just an approximation and does not apply to meshed network. Then, whenever the OLM routine identifies a limit violation, the tree to which the binding constraint pertains is detected and the FAR is applied only to the set of loads that belongs to the tree (see Figure 5). This simplification is justified by the fact that loads outside the tree are expected to have very small sensitivities and thus, are not going to be rescheduled by the solution of problem (1) or (2).

D. Modeling of Electric Vehicle Chargers

In the proposed optimization problems (1) and (2), only the active power p_i of loads has been considered. However, when solving the power flow problem, a detailed steady state model of all loads is used. In our formulation, we have used the well-known Voltage Dependent Load (VDL) model, i.e., load powers are monomial functions of the bus voltage magnitude, as follows:

$$p_i = p_i^0 (v_i/v_i^0)^{\alpha_{p,i}} \qquad (14)$$
$$q_i = q_i^0 (v_i/v_i^0)^{\alpha_{q,i}}$$

where v_i^0 is the rated voltage at the load bus. For example, a constant shunt admittance can be modeled as a VDL for which $\alpha_{p,i} = \alpha_{q,i} = 2$ and:

$$g_i^0 = \frac{p_i^0}{(v_i^0)^2}, \qquad b_i^0 = \frac{q_i^0}{(v_i^0)^2} \qquad (15)$$

The model of electric vehicles depends on the charger and on the power electronic system. In our simulations we have considered that for electric vehicle chargers, $\alpha_{p,i} = 1$, i.e., a constant current model and $q_i^0 = 0$, which leads to:

$$p_i = p_i^0 (v_i/v_i^0) \qquad (16)$$
$$q_i = 0$$

E. Inclusion of Distributed Energy Resources

From the OLM viewpoint, DERs are similar to loads since they can offer a certain amount of power during a certain period. As said above, DER admissions undergo same rules as loads, except for the FAR which does not apply to DER

offers. However, the sudden outage of some DERs can cause a blackout if the load shedding is not fast enough, especially if the penetration of DERs in the grid is high.

Since an accurate contingency analysis cannot be solved on-line, the OLM can only implement approximated security analysis techniques. With this aim, security current limits $i_s^{\max}$ and security voltage limits $v_s^{\max}$ and $v_s^{\min}$ are imposed only to define the admission of DER power offers and are computed as follows:

$$i_s^{\max} = (1 - s)\, i^{\max} \qquad (17)$$
$$v_s^{\max} = (1 - s)\, v^{\max}$$
$$v_s^{\min} = (1 + s)\, v^{\min}$$

where $i^{\max}$, $v^{\max}$ and $v^{\min}$ are technical, thermal and voltage limits and s is the "security margin". The value of s depends on the system operator and should be based on the knowledge of the system. A reasonable range for s is $[0.05, 0.2]$.

III. Case Study

The simulation is obtained considering the 50-bus radial network shown in Fig. 6 and a time horizon of 24 h, i.e., 86400 s, with a step length of 10 s, which leads to a total of 8640 steps. The network is modeled as a single-phase equivalent and loads can be connected to any bus. Two subnetworks modeled as radial unbalanced three-phase systems are connected to buses 39 and 42. These sub-networks are not represented in Fig. 6. The interface between the single-phase equivalent and the three-phase system is modeled as proposed in [19]. In this case study, line thermal limits are considered "small" in order to force congestion and "stress" the OLM routine with relatively big load requests queues.
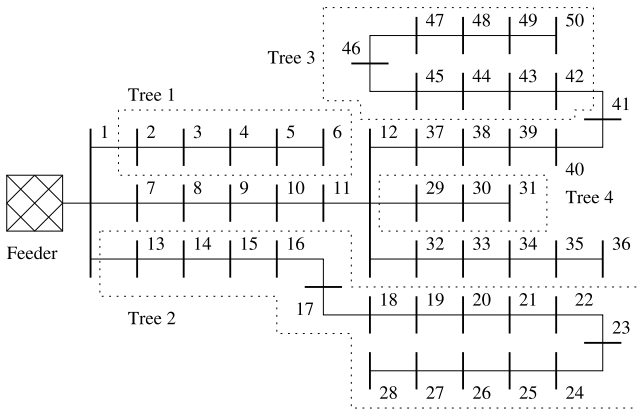

Fig. 6. 50-bus radial system.

To compare different settings two indices are defined: (i) the total supplied energy (TSE) in kWh of electric vehicle chargers; and (ii) the cumulative number of supplied electric vehicles (NSE). By definition, both TSE and NSE are zero at the beginning of the simulation. The absolute values of such indices is not relevant *per se*, rather their relative values provide a quantitative indication of how efficient and how fair a particular set of OLM setting parameters is in comparison to another set.

TABLE I
PERFORMANCE OF THE FAR FOR THE 50-BUS RADIAL NETWORK.

Solver	t_{tot} [s]	t_{mean} [ms]
No FAR	15.56	1.80
FAR with (1)	50.57	5.85
FAR with (2)	17.96	2.08

The computational burden is defined in terms of the CPU time required to complete a simulation. In the next sections, t_{tot} indicates the total time required to solve the OLM routine for the simulated 24 h, whereas t_{mean} indicates the average time required for each snapshot, i.e., $t_{\text{mean}} = t_{\text{tot}}/8640$. It has to be noted that the variance of the time needed to solve each snapshot can highly vary since it depends on how many new requests have to be handled, whether congestion occurs, whether sensitivities has to be updated, etc.

Load requests are generated using a random number generator. To properly compare results, the seed that initializes the random number generator is the same for all simulations. However, depending on the settings of the OLM, the number of calls to the random number generators can vary and, hence, also the sequence of load requests can vary.

All simulations were solved on a 64-bit Linux operating system running on an Intel i7 2.67 GHz processor with 8 GB of RAM. The OLM algorithms has been implemented in Dome[1] and the power flow is solved using a backward-forward sweep method as described in [20].

A. Impact of the Fair Admission Rule

This section shows a comparison of the impact of the fair admission rule on the performance indices TSE and NSE. Three scenarios are considered: (i) no FAR enforced; (ii) FAR based on the solution of the full QP problem (1); and (iii) FAR based on the solution of the iterative approximated QP problem (2). In this case study, sensitivities are computed for all loads (i.e., without enforcing location reduction) and updated every 10 new admitted load requests.

Table I shows the CPU times for the three considered cases. The main conclusions that can be drawn by these simulations are twofold: (i) the computational burden of computing sensitivities is relatively low; and (ii) the computational burden if the solution of the full QP problem (1) is significantly higher than that of the approximated iterative QP problem (2).

Figure 7 compares the TSE and NSE indices for the three different FAR methods discussed above. As expected, the poorest performance is given by the OLM without enforcing the FAR. Then, the approximated QP problem (2) performs worse than the full QP problem (1). This result was also to be expected, as the optimization problem (1) provides an optimal solution, while (1) only a sub-optimal one. However, it has to be noted that (1) is intrinsically slower than (2), and such difference increases more then linearly as the size of the system increases. In fact, the computational complexity of (2) is close to be linear, while the computational complexity of the

---

[1]Available at `faraday1.ucd.ie/dome.html`

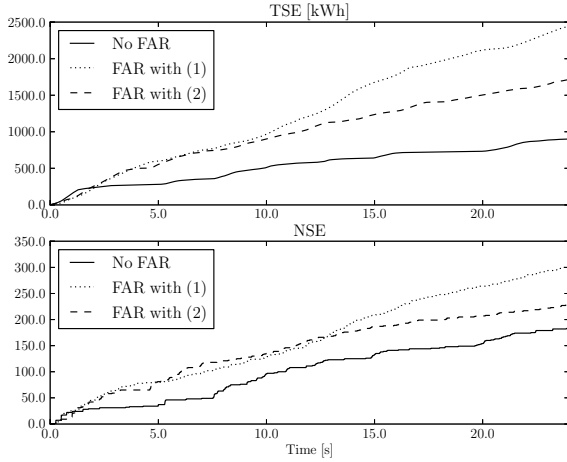(1) is polynomial and similar to that of the LU factorization algorithm (which is required to solve QP problems).

Fig. 7. Comparison of TSE and NSE indices for different fair admission rule methods for the 50-bus radial system.
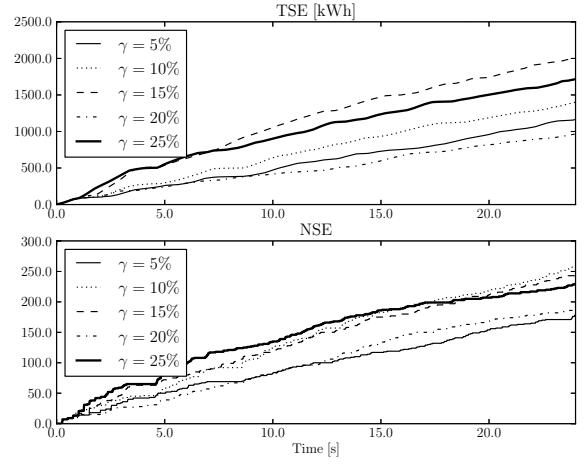

Fig. 8. Comparison of TSE and NSE indices for different values of the power reduction threshold γ for the 50-bus radial system.

TABLE II
PERFORMANCE OF THE OLM ROUTINE FOR DIFFERENT VALUES OF m_{new} FOR THE 50-BUS RADIAL NETWORK.

m_{new}	t_{tot} [s]	t_{mean} [ms]
5	12.46	1.44
10	17.96	2.08
15	17.35	2.01
30	12.05	1.39

B. Impact of Power Reduction Threshold

This section shows a comparison of the impact of the power reduction threshold γ on OLM performance and on indices TSE and NSE. The considered values for the coefficient γ are 5%, 10% 15%, 20%, and 25%. For these simulations, FAR is solved using the iterative approximated QP problem (2). Moreover, sensitivities are computed for all loads (i.e., without enforcing location reduction) and updated every 10 new admitted load requests.

Figure 8 shows the CPU times and the TSE and NSE indices for the five considered values of γ. The main conclusions that can be drawn by these simulations are threefold: (i) the computational cost is not strictly related to the value of γ, unless γ is very small (see the case $\gamma = 5\%$); (ii) the performance and thus TSE and NSE indices of the OLM routine depends on the value of γ, however, it is not possible *a priori* to define the optimal value of γ; and (iii) there is no strict correspondence between the TSE and the NSE indices as γ varies.

As it can be observed in Figure 8, the highest values of the TSE index are obtained for $\gamma = 15\%$, while the worst case is obtained for $\gamma = 20\%$. However, if observing the NSE index, one can only say that $\gamma = 10\%$, 15% or 25% are better choices than $\gamma = 5\%$ or 20%. It has also to be noted that a value of γ can be optimal for a certain system and for a certain set of other OLM setting parameters, but less good for another network or another set of parameters. Hence, the value of γ, which is actually arbitrary, should be tuned based on historical data of the network, or, using a trial-and-error approach. In the remainder of this section, $\gamma = 25\%$ is assumed, since it provides overall good results for all OLM parameter set.

C. Effect of Sensitivity Time Reduction

This section discusses the effect of sensitivity time reduction on the performance of the OLM routine. As discussed in Subsection II-C, time reduction consists in updating the sensitivities not at each step of the OLM routine, but with a lower frequency. With this aim, let m_{new} the threshold number of new load admissions that have to occur before sensitivities are updated. Table II and Figure 9 show the CPU times and TSE ans NSE indices for four value of m_{new}, namely 5, 10, 15 and 30. The FAR used for these simulations is that based on the solution of problem (2).

The calculation of sensitivities is not strictly related to the computational burden and to OLM accuracy, at least for values of $m_{\text{new}} \leq 15$. In other words, the number of sensitivities updates has generally little impact on the total simulation time. On the other hand, if sensitivities are updated with a reasonable frequency, the OLM performance measured in terms of TSE and NSE indices is similar. This fact basically indicates that sensitivities change little and that their impact on the solution of problem (2) is not really crucial. In the remainder of this section, $m_{\text{new}} = 10$ will be used.

D. Effect of Sensitivity Location Reduction

This section discusses the effect of sensitivity location reduction on the performance of the OLM routine. As discussed in Subsection II-C, location reduction consists in defining sections of the network, called *trees*, that are assumed to be
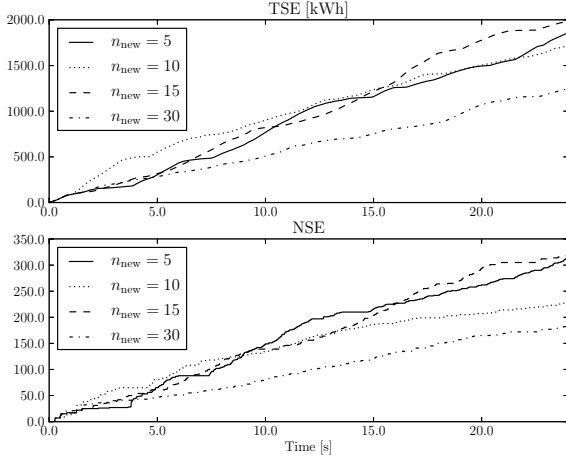
Fig. 9.   Comparison of TSE and NSE indices for different values of the parameter m_{new} for the 50-bus radial system.

Fig. 10.   Comparison of TSE and NSE indices for the OLM routine with and without location reduction for the 50-bus radial system.

TABLE III

PERFORMANCE OF THE OLM ROUTINE WITH AND WITHOUT LOCATION
REDUCTION FOR THE 50-BUS RADIAL NETWORK.

Method	FAR type	t_{tot} [s]	t_{mean} [ms]
Full network	Problem (1)	50.57	5.85
Trees	Problem (1)	147.87	17.11
Full network	Problem (2)	17.96	2.08
Trees	Problem (2)	13.90	1.61

TABLE IV

PERFORMANCE OF THE OLM ROUTINE FOR WITH AND WITHOUT
UPDATING LOAD REQUESTS AND FOR THE 50-BUS RADIAL NETWORK.

Method	t_{tot} [s]	t_{mean} [ms]
No request update	17.96	2.08
Request update	71.40	8.26

decoupled in what concerns sensitivities. With this aim, 4 trees are defined for the 50-bus system (see Fig. 6).

Table III and Figure 10 show the CPU times and TSE and NSE indices for the 50-bus system with and without defining trees. Both full QP problem (1) and approximated QP problem (2) are compared.

Relevant remarks are the following.

1) Location reduction allows reducing the computational burden if using the approximated QP problem (2). In case of using trees with the full QP problem (1), the simulation time increases drastically.
2) Reducing the computational burden can be obtained at the cost of reducing the performance of the OLM routine, e.g., reducing both TSE and NSE indices.
3) The differences between the OLM routine using the full sensitivity set and the OLM routine using location reduction is due to the fact that network branches are actually not fully decoupled.

E. Effect of Automatic Load Rescheduling

So far, it has been assumed that the OLM routine can reschedule power requests of EV chargers and that, once rescheduled, EV chargers do not attempt to reformulate the original power request. Actually, it is to be expected that EV chargers continuously update their request based on the current supplied power.

Table IV and Figure 11 show the CPU times and TSE and NSE indices for the 50-bus system with and without load
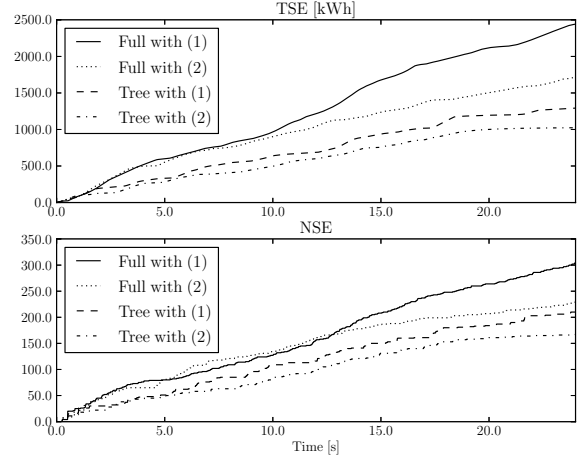
request update. Simulations are solved using the approximated QP problem (2). As it can be observed, updating load requests drastically slows down the OLM routine because if forces reconsidering the same load request several times until it is either accepted at the original requested power or the total request energy is supplied. On the other hand, updating load requests allows supplying more power to admitted loads, even if accepting a reduced number of new requests.

F. Effect of DERs and Security Limits

This section studies the effect of the security margin s on the admission control of DERs. The 50-bus system is modified to include 5 DER devices at buses 5, 22, 36, 41, and 45. Different levels of security margins s are considered, namely 0, 5%, 10% and 20%. Refer to Subsection II-E for the details on how the security margin modifies technical limits.

Figure 12 shows the CPU times and TSE and NSE indices for the 50-bus system with inclusion of DERs and for different values of the security margins. Simulations are solved using the approximated QP problem (2). DER power offers are generated randomly with a mechanism similar to load power requests. The volatility of random generation power offers is much higher than that that would happen in real-world systems. However, this is done by purpose to force the OLM algorithm to accommodate high volatility of both loads and DERs. Table V shows that using a too conservative value for s can lead to longer execution times. This result is due to the fact that higher values of s lead to bigger queues. Figure 12 indicates that the higher the security margin s, the lower the TSE and NSE indices. Again, this result was to be expected as
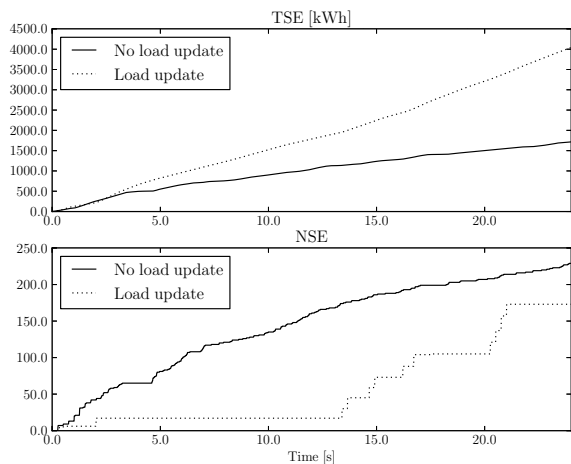
Fig. 11. Comparison of TSE and NSE indices for the OLM routine with and without updating load requests for the 50-bus radial system.

Fig. 12. Comparison of TSE and NSE indices for the 50-bus radial system with DERs and different values of the security margin s.

the higher the security margin s, the more likely DER power offers are rejected and put in the queue.

TABLE V

PERFORMANCE OF THE OLM ROUTINE WITH DERS AND DIFFERENT VALUES OF THE SECURITY MARGIN s FOR THE 50-BUS RADIAL NETWORK.

Security margin s	t_{tot} [s]	t_{mean} [ms]
0	22.25	2.58
5%	16.20	1.88
10%	20.08	2.32
20%	26.67	3.09

IV. CONCLUSIONS

This paper proposes two optimization problems for load demand management aimed to increase served load while preserving system security. The proposed methods can also handle distributed energy resources. The version of the fair admission rule based on a full quadratic programming optimization problem provides performance indices about 50% higher than those obtained with the admission control strategy version based on an iterative approximated quadratic programming problem. Since a higher accuracy is obtained at a cost of a higher computational burden, the choice of the optimization problem depends on the size of the problem.

Future work will focus on further reduce the computational burden through adequate implementation of the proposed technique. The authors are currently working on a production-grade implementation of the proposed technique.

REFERENCES

[1] S. Deilami, A. S. Masoum, P. S. Moses, and M. A. S. Masoum, "Real-Time Coordination of Plug-In Electric Vehicle Charging in Smart Grids to Minimize Power Losses and Improve Voltage Profile," *IEEE Transactions on Smart Grid*, vol. 2, no. 3, pp. 456–467, Sept. 2011.

[2] A. S. Masoum, S. Deilami, P. S. Moses, M. A. S. Masoum, and A. Abu-Siada, "Smart Load Management of Plug-in Electric Vehicles in Distribution and Residential Networks with Charging Stations for Peak Shaving and Loss Minimization considering Voltage Regulation," *IET Proceedings on Generation, Transmission and Distribution*, vol. 5, no. 8, pp. 877–888, Sept. 2011.
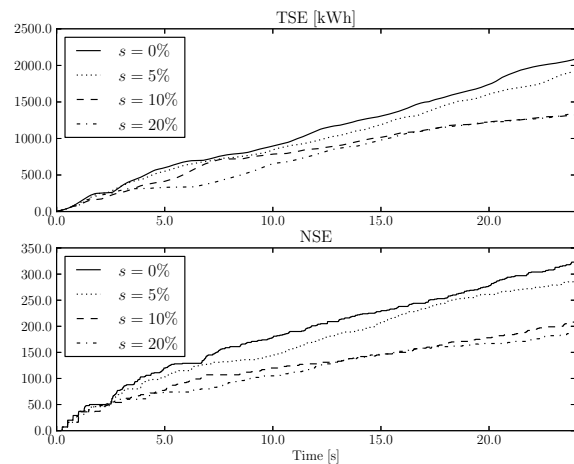
[3] A. J. Conejo, J. M. Morales, and L. Baringo, "Real-Time Demand Response Model," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 236–242, 2010.

[4] S. Bruno, S. Lamonaca, M. La Scala, G. Rotondo, and U. Stecchi, "Loac Control through Smart-metering on Distribution Networks," in *Proceedings of the IEEE Power Tech Conference*, Bucharest, Romania, June 2009.

[5] H. S. V. S. Kumar Nunna and S. Doolla, "Demand Response in Smart Distribution System With Multiple Microgrids," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1678–1687, Dec. 2012.

[6] Z. Fan, "A Distributed Demand Response Algorithm and Its Application to PHEV Charging in Smart Grids," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1280–1290, Sept. 2012.

[7] E. Sortomme and A. A. El-Sharkawi, "Optimal Charging Strategies for Unidirectional Vehicle-to-Grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 131–138, Mar. 2011.

[8] W. Kempton and J. Tomic, "Vehicle-to-grid Power Implementation: From Stabilizing the Grid to Supporting Large-scale Renewable Energy," *J. Power Sources*, vol. 144, no. 1, pp. 280–294, 2005.

[9] J. R. Pillai and B. Bak-Jensen, "Impacts of Electric Vehicle Loads on Power Distribution Systems," in *Proceedings of the IEEE Vehicle Power and Propulsion Conference (VPPC)*, Lille, France, Sept. 2010.

[10] Z. Darbai and M. Ferdowsi, "Aggregated Impact of Plug-in Hybrid Electric Vehicles on Electricity Demand Profile," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 4, pp. 501–508, Oct. 2011.

[11] S. Shao, M. Pipattanasomporn, and S. Rahman, "Grid Integration of Electric Vehicles and Demand Response With Customer Choice," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 543–550, Mar. 2012.

[12] S. Bae and A. Kwasinski, "Dynamic Modeling and Operation Strategy for a Microgrid With Wind and Photovoltaic Resources," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1867–1876, Dec. 2012.

[13] A. M. Leite da Silva, L. C. Nascimento, M. A. da Rosa, D. Issicaba, and J. Peças Lopes, "Distributed Energy Resources Impact on Distribution System Reliability Under Load Transfer Restrictions," *IEEE Transactions on Smart Grid*, pp. 1–9, 2012, preprint.

[14] J. R. Pillai and B. Bak-Jensen, "Integration of Vehicle-to-Grid in the Western Danish Power System," *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 12–19, Mar. 2011.

[15] R. Minguez, F. Milano, R. Zárate-Miñano, and A. J. Conejo, "Optimal Network Placement of SVC Devices," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1851–1860, Nov. 2007.

[16] I. E. Grossmann, J. Viswanathan, A. Vecchietti, R. Raman, and E. Kalvelagen, *GAMS/DICOPT: A Discrete Continous Optimization Package*, Engineering Research Design Center, Carnegie Mellon University, Pittsburg, PA, 2002, available at http://www.gams.com/.

[17] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming, Theory and Algorithms, 2nd ed.* New York, USA: Wiley, 1993.

[18] A. Conejo, E. Castillo, R. Mínguez, and R. García-Bertrand, *Decom-*

position Techniques in Mathematical Programming. Engineering and Science Applications. New York, USA: Springer, 2006.

[19] J. M. T. Marinho and G. N. Taranto, "A Hybrid Three-Phase Single-Phase Power Flow Formulation," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1063–1070, Aug. 2008.

[20] D. Shirmohammadi, H. Hong, A. Semlyen, and G. Luo, "A Compensation-based Power Flow Method for Weakly Meshed Distribution and Transmission Networks," *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 753–762, May 1988.

Federico Milano (S'09) received from the University of Genoa, Italy, the Electrical Engineering degree and the Ph.D. degree in 1999 and 2003, respectively. From 2001 to 2002 he was a Visiting Scholar at the University of Waterloo, Canada. From 2003 to 2013, he was with the University of Castilla-La Mancha, Spain. In 2013, he joined the University College Dublin, Ireland, where he is currently Associate Professor. His current research interests include power system modeling, stability and control.

Olivier Hersent is a graduate of École Polytechnique and Telecom Paris. In 1998 he founded Net-Centrex (VoIP softswitch systems). In 2006 he became CTO of Comverse, inc (CMVT). He founded Actility in 2010, a company specialized in large scale automation systems and demand-response. Olivier Hersent was awarded the SEE Andre Blondel award in 2012.