This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2021.3137842, IEEE Transactions on Power Systems

A novel optimization approach for sub-hourly unit commitment with large numbers of units and virtual transactions

Jianghua Wu, *Student Member, IEEE*, Peter B. Luh, *Life Fellow, IEEE*, Yonghong Chen, *Senior Member, IEEE,* Mikhail A. Bragin, *Member, IEEE,* Bing Yan, *Member, IEEE*

Abstract **– Unit Commitment (UC) is an important problem in power system operations. It is traditionally planned for 24 hours with one-hour time intervals. To accommodate the increasing net-load variability, sub-hourly UC has been suggested for improved system flexibility. Such a problem is larger and more complicated than hourly UC because of the increased number of periods and reduced unit ramping capabilities per period. The computational burden is further exacerbated for systems with large numbers of virtual transactions leading to dense transmission constraint matrices. Consequently, the state-of-the-art and practice method, branch-and-cut (B&C), suffers from poor performance. In this paper, our recent Surrogate Absolute-Value Lagrangian Relaxation (SAVLR) is enhanced by embedding ordinal-optimization concepts for a drastic reduction in subproblem solving time. Rather than formally solving subproblems by using B&C, subproblem solutions satisfying SAVLR's convergence condition are obtained by modifying solutions from previous iterations or solving crude subproblems. All virtual transactions are included in each subproblem to reduce major changes in solutions across iterations. A parallel version is also developed to further reduce the computation time. Testing on MISO's large cases demonstrates that our ordinal-optimization embedded approach obtains near-optimal solutions efficiently, is robust, and provides a new way of solving other MILP problems.**

Index Terms **-- Ordinal Optimization, Parallel Processing, Surrogate Absolute-Value Lagrangian Relaxation, Sub-hourly Unit Commitment**

NOMENCLATURE

Index

i	index for conventional units
v	index for virtual transactions
y	index for dispatchable demand bids
n	index for nodes
z	index for three types of reserves
l	index for transmission lines
t	Index for time periods
s	index for three types of start-ups
j	index for subproblems

Parameters

$D_{n,t}$	system demand at node n at time t
$R_{z,t}$	required amount for reserve type z at time t
$\overline{F}_l$, $\underline{F}_l$	maximum and minimum transmission capacities of line l
$\alpha_{n,l}$	generation shift factor at node n for line l
T_i^{MU}	minimum up time of unit i
T_i^{MD}	minimum down time of unit i
$\overline{P}_{i,t}$, $\underline{P}_{i,t}$	maximum and minimum power output for unit i at time t
RR_i	ramp rate for unit i
TE_i	maximum daily energy of unit i
TS_i	maximum daily start-up times of unit i
$\overline{M}_{v,t}$, $\underline{M}_{v,t}$	minimum and maximum generations (or demands) of virtual transaction v at time t
$\overline{D}_{y,t}$	maximum level of dispatchable demand bid y at time t
$C_{i,s,t}^{Start}$	cost coefficient for start-up s of unit i at time t
$C_{i,t}^{NL}$	no-load cost coefficient of unit i at time t
C^E	piece-wise linear generation cost of unit i
$C_{i,z,t}^R$	cost coefficient for reserve z of unit i at time t
$C_{v,t}^V$	cost coefficient of virtual transactions v at time t
$C_{y,t}^Y$	cost coefficient of dispatchable demand bid y at time t
c^T	penalty coefficient of transmission capacity constraints' violations
c^D	penalty coefficient of system demand constraints' violations
c^R	penalty coefficient of system reserve constraints' violations

Decision Variables

$x_{i,t}$	commitment status of unit i at time t
$u_{i,t}$	start-up status of unit i at time t
$w_{i,t}$	shut-down status of unit i at time t
$b_{i,s,t}$	three types (hot, intermediate and cold) of start-ups
$p_{i,t}$	generation level of unit i at time t
$m_{v,t}$	generation level of virtual transaction v at t
$d_{y,t}$	power required by dispatchable demand bid y at time t
$r_{i,z,t}$	reserve contribution for type z of unit i at time t
$f_{t,l}$	power flow through line l at time t
λ_t	Lagrangian multipliers for system demand constraints
$\overline{s}_{t,l}$, $\underline{s}_{t,l}$	non-negative slack variables for transmission constraints
$s_{z,t}^R$	slack variables for system-wide reserve constraints
$q_t^{D,+}$, $q_t^{D,-}$	non-negative variables for linearization

I. INTRODUCTION

UNIT Commitment (UC) is an important problem in power system operations – it identifies how to meet the system demand by committing units and deciding generation levels while minimizing the total cost of production subject to individual unit constraints and system-wide reserve and transmission capacity constraints. A UC problem is generally formulated as a Mixed-Integer Linear Programming (MILP) problem over a 24-hour horizon with one hour as the time interval. Increasing dynamics on the grid prompted the industry to consider whether UC with sub-hourly intervals would increase system performance [1]. Sub-hourly UC has thus been suggested as a way to improve system flexibility and reliability

This work is supported in part by the Midcontinent ISO and by the National Science Foundation under grants ECCS-1810108 and ECCS-1831811. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the MISO and NSF.

J. Wu, P. B. Luh, and M. A. Bragin are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, 06269 (e-mails: jianghua.wu; peter.luh; mikhail.bragin @uconn.edu).

Y. Chen is with MISO, Carmel, IN, 46032 (e-mails: ychen@misoenergy.org).

B. Yan is with the Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology, Rochester, NY, 14623 (e-mail: bxyeee@rit.edu).

because it can accommodate greater intra-hour net load variability [2]-[7]. Sub-hourly UC, however, is much more complex than hourly UC because of 1) the increased number of periods leading to larger problem sizes, and 2) the much reduced unit ramping capabilities per period resulting in more complicated convex hulls (the smallest convex set enclosing all feasible solutions) as presented in Figure 9 of [8]. The computational burden is further exacerbated for systems with a large number of virtual transactions and dispatchable demand bids which cause dense transmission capacity constraint matrices [9]. As a result, even without explicitly modeling of uncertainties caused by, e.g., intermittent renewables, deterministic sub-hourly UC is difficult to solve. In this paper, deterministic sub-hourly UC with large numbers of units, virtual transactions and dispatchable demand bids is considered with 15 minutes as the time interval, and the goal is to obtain near-optimal solutions within 30 minutes.

As will be reviewed in Section II, branch-and-cut (B&C) [10], the state-of-the-art and practice method for hourly UC, cannot handle the increased complexity and suffers from poor performance for sub-hourly UC. Lagrangian Relaxation (LR) [11][12] was one of the earlier methods for hourly UC. It reduces complexity by relaxing coupling constraints and decomposing the relaxed problem into subproblems. Standard LR, however, has several major difficulties, such as high computational requirements, zigzagging of multipliers, and the need to adaptively guesstimate the unknown optimal dual value. Surrogate Lagrangian Relaxation (SLR) overcame these major difficulties [13]. Its convergence has then been accelerated by adding absolute-value penalty terms in our recent Surrogate Absolute-Value Lagrangian Relaxation (SAVLR) method [14]. Within SAVLR, MILP subproblems are normally solved by using B&C. Subproblem solving, however, may still be time-consuming for sub-hourly UC with large numbers of units, virtual transactions and dispatchable demand bids.

In Section III, the deterministic sub-hourly UC formulation, which is the same as that of hourly UC but with 15 minutes as the time interval, is briefly presented. System demand constraints and reserve requirements should be strictly satisfied. Transmission capacity constraints, however, are modeled as "soft" and allowed to be violated with a predetermined penalty coefficient. Additionally, although uncertainties are not explicitly modeled, three types of reserves are included. This is the current standard practice for Independent System Operators (ISOs) to manage uncertainties.

In Section IV, our solution methodology is presented. To avoid introducing too many multipliers, only system demand constraints are relaxed. System reserve constraints are converted to soft constraints following the approach of [15], which, together with soft transmission capacity constraints, are not relaxed. More importantly, inspired by the Ordinal Optimization (OO) concepts that an "order" is easier to obtain than "values" and a problem with a softened goal is easier to solve than the original problem, a novel approach is developed by embedding the OO concepts within SAVLR. Specifically, rather than formally solving a subproblem by using B&C, "good enough" feasible subproblem solutions that satisfy the SAVLR convergence condition are obtained by modifying solutions from previous iterations or solving crude subproblems following OO concepts [16][17]. B&C is called to solve a subproblem only when such a good-enough solution cannot be obtained. This novel idea leads to a drastic reduction in CPU times because B&C is rarely called. Virtual transactions and dispatchable demand bids are included in all subproblems to reduce multiplier zigzagging and improve convergence. Finally, a parallel version is also developed to further reduce the CPU time.

In Section V, multiple Midcontinent ISO (MISO)'s "hard" cases, whose solutions are difficult to obtain within 20 or 30 minutes even for hourly UC by using B&C [9], are tested with 15 minutes as the time interval over a horizon of 36 hours. Results demonstrate that our approach obtains high-quality solutions in a computationally efficient way, significantly outperforms existing methods, and is robust.

This manuscript is a major improvement over our preliminary results presented at the 2020 IEEE Power and Energy Society General Meeting [18]. Key enhancements include: 1) the specific use of OO concepts in SAVLR for sub-hourly UC is elaborated; 2) a parallel version is developed to further reduce the CPU time; 3) more MISO's hard cases are tested and analyzed to demonstrate the performance and robustness of our method; and 4) the reasons why B&C is rarely used to solve subproblems are examined. Our approach presents a new optimization concept to solve subproblems by not using standard MILP methods, and lead to significant reduction of computational requirements. It will have vital implications on solving other complex MILP problems in power systems and beyond.

II. Literature Review

Subsection II.A reviews the literature on sub-hourly UC. In subsection II.B, branch-and-cut (B&C), the standard method to solve hourly UC, is presented. Decomposition and coordination approaches based on Lagrangian Relaxation are reviewed in subsection II.C. In subsection II.D, the Ordinal Optimization (OO) concepts are presented.

A. Sub-hourly unit commitment

With the increasing dynamics on the grid, hourly UC cannot capture the sub-hourly net load variability [1]. Sub-hourly UC has thus been suggested as a way to improve system flexibility and reliability [2]. In [3] and [4], deterministic sub-hourly UC is compared with hourly UC. With the increased number of time intervals, sub-hourly UC captures more variability in system demand, leading to more economic solutions than hourly UC. In [5], both deterministic and stochastic sub-hourly UC are compared with hourly UC. It is shown that sub-hourly dispatch results have lower costs. In [6], reserves are shown to be significantly lowered for sub-hourly UC than hourly UC under high penetration of intermittent renewables. The impacts of sub-hourly UC on power system dynamics are analyzed in [7]. It was shown that long-term frequency deviation is reduced for sub-hourly UC, leading to improved reliability. However, in view that sub-hourly UC is much more difficult to solve than hourly UC as explained earlier, it is mostly used in near real-time markets looking ahead 1-3 hours. It is relatively new to apply it in day-ahead markets with a horizon of 24-36 hours.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2021.3137842, IEEE Transactions on Power Systems

B. Branch-and-cut (B&C)

UC problems are generally formulated as MILP problems, and solved by using B&C [10]. For a given problem, the method applies "valid cuts" and tries to delineate the convex hull of feasible solutions. If the convex hull or its facets adjacent to the optimal solution are obtained, then the optimal solution can be quickly obtained by solving the corresponding Linear Programming (LP) problem. If the above is difficult to achieve as explained in Section I, then the method relies on time-consuming branch-and-bound. In [3] and [6], B&C is used to solve small sub-hourly UC problems with less than 100 units. For large sub-hourly UC problems, however, B&C suffers from poor performance when directly applied. In [19], solutions of MISO hourly UC from B&C are used to provide initial solutions for sub-hourly UC. Good solutions are then obtained by a "polishing" method, which adaptively fixes binary and continuous variables while filtering out the constraints that are unlikely to be violated.

C. Decomposition and coordination approaches

Lagrangian Relaxation (LR) is a price-based decomposition and coordination method, and was one of the early methods to solve UC problems [12]. It reduces complexity by relaxing coupling constraints and decomposing the relaxed problem into subproblems, which are coordinated by iteratively updating Lagrangian multipliers based on subgradient directions. The standard LR methods, however, have several major difficulties: 1) significant efforts to obtain a subgradient – requiring solving all subproblems optimally; 2) zigzagging of multipliers in view of the geometry of the dual function for MILP problems; and 3) the need to guesstimate the unknown optimal dual value.

All the major difficulties mentioned above have recently been overcome in the Surrogate Lagrangian relaxation (SLR) method [13]. SLR updates Lagrangian multipliers based on "surrogate" subgradients [20], which are obtained by solving one or a few subproblems not to optimality, but as long as the "surrogate optimality condition" (see (29) in subsection IV.B) is satisfied. Since only a subset of subproblems needs to be solved to update multipliers, the computational requirements are much reduced; and the changing of surrogate subgradient directions across iterations is also reduced as compared to that of the traditional LR, leading to much smoothened multiplier trajectories. Moreover, unlike previous LR-based methods, SLR does not require the knowledge of the unknown optimal dual value for convergence proof as well as for practical implementations. Recently, the convergence of SLR has been significantly improved by introducing absolute-value terms, which are exactly linearizable, to penalize the violations of relaxed system-wide constraints in the Surrogate Absolute-Value Lagrangian Relaxation (SAVLR) method [14]. Subproblems in SLR and SAVLR are generally solved as MILP problems by using B&C. This, however, may still take a long time for sub-hourly UC problems with large numbers of units, virtual transactions and dispatchable demand bids.

D. Ordinal Optimization

Ordinal Optimization (OO) has been effectively used in computationally intensive simulation-based optimization, and has two major concepts [17]. The first is that an "order" is easier to obtain than "values." Taking two objects A and B as an example, it is easier to know which object is heavier than to know the exact weights of A and B. Second, a problem with a softened goal is easier to solve than the original problem. For example, it is easier to obtain a solution that falls within top the 5% of all solutions than to obtain the optimal solution. OO thus uses crude models and quick simulation runs to roughly order solution candidates, and then select solutions that are good enough with high probabilities for further exploration. OO has recently been used to solve subproblems in generalized assignment problems [21].

III. Problem Formulation

This section considers a power system with I conventional units, V virtual transactions, Y dispatchable demand bids, N nodes, Z types of reserves, and L transmission lines, which are distributed among J areas. The 15-minute UC is formulated as an MILP problem following [22]. The formulation is the same as that for hourly UC, except that 15 minutes are used as the time interval over T periods (or $T/4$ hours). Constraints include (1) system-wide demand, reserve, and transmission capacity constraints; (2) individual unit-level constraints, e.g., generation capacity and ramp-rate constraints for conventional units; and capacity constraints for virtual transactions and dispatchable demand bids as presented below.

<u>*Constraints*</u>

System Demand Constraints. The total generation from all resources should equal system demand at each period, i.e.,

$$\sum_{i=1}^{I} p_{i,t} + \sum_{v=1}^{V} m_{v,t} - \sum_{y=1}^{Y} d_{y,t} = \sum_{n=1}^{N} D_{n,t}, \forall t, \quad (1)$$

where the continuous generation level of unit i $(1 \leq i \leq I)$ at time t $(1 \leq t \leq T)$ is denoted as $p_{i,t}$, the continuous generation level of virtual transaction v $(1 \leq v \leq V)$ at t is denoted as $m_{v,t}$, and the continuous power required by dispatchable demand bid y $(1 \leq y \leq Y)$ at t is denoted as $d_{y,t}$. The system demand at node n $(1 \leq n \leq N)$ at t is denoted as $D_{n,t}$.

System Reserve Constraints. In the current standard practice for ISOs, to maintain reliability, reserves are used to manage uncertainties. Following [23], three types of reserves including regulation, regulation plus spinning, and operating reserve are considered, indexed by z = 1, 2, and 3, respectively:

$$\sum_{i=1}^{I} r_{i,z,t} \geq R_{z,t}, \forall z, \forall t, \quad (2)$$

where the amount of reserve contribution of unit i at time t for type z of reserve is denoted as the continuous variable $r_{i,z,t}$, and the required amount of type z reserve at time t is denoted as $R_{z,t}$.

Transmission Capacity Constraints. DC power flow is considered, and the flow in line l $(1 \leq l \leq L)$ at t, $f_{t,l}$, cannot exceed the line's capacities at each period:

$$f_{t,l} - \overline{s}_{t,l} \leq \overline{F}_l, \forall t, \forall l, \quad (3)$$

$$f_{t,l} + \underline{s}_{t,l} \geq \underline{F}_l, \forall t, \forall l, \text{ with} \quad (4)$$

$$f_{t,l} = \sum_{n=1}^{N} \alpha_{n,l} \left(\sum_{i \in I_n} p_{i,t} + \sum_{v \in V_n} m_{v,t} - \sum_{y \in Y_n} d_{y,t} - D_{n,t} \right), \forall t, \forall l. \quad (5)$$

In the above, transmission capacities of line l are denoted as $\underline{F}_l$ and $\overline{F}_l$; the sets of units, virtual transactions and dispatchable demand bids at node n are denoted as I_n, V_n, and Y_n, respectively. The generation shift factor $\alpha_{n,l}$ indicates the change of power flow through line l with respect to a change in injection at node n. Following [15], the above transmission capacity constraints (3) and (4) are modeled as "soft" constraints, and are allowed to be violated by non-negative continuous variables $\overline{s}_{t,l}$ and $\underline{s}_{t,l}$ with a fixed penalty coefficient c^T as will be seen in (18).

Individual unit-level constraints. Minimum up/down-time constraints follow Equations (6)-(7) of [22]:

$$\sum_{\tau=t-MU_i}^{t} u_{i,\tau} \le u_{i,t}, t=[T_i^{MU},T], \forall i, \tag{6}$$

$$\sum_{\tau=t-MD_i}^{t} w_{i,\tau} \le 1-u_{i,t}, t=[T_i^{MD},T], \forall i, \tag{7}$$

where start-up and shut-down statuses for unit i at time t are denoted as binary variables $u_{i,t}$ and $w_{i,t}$, respectively; and minimum up and down times are denoted as T_i^{MU} and T_i^{MD}, respectively. The following logical constraints guarantee that $u_{i,t}$ and $w_{i,t}$ take the appropriate values when unit i starts up or shuts down:

$$x_{i,t} - x_{i,t-1} = u_{i,t} - w_{i,t}, \forall i, \forall t, \tag{8}$$

where commitment status for unit i at time t is denoted as the binary variable $x_{i,t}$. Capacity constraints, ramping constraints, and reserve limits are given in (A9)-(A10), (A11) and (A12) of [24], respectively. They are described as follows:

$$p_{i,t} + \sum_{z=1}^{3} r_{i,z,t} \le \overline{P}_{i,t} x_{i,t}, \forall i, \forall t, \tag{9}$$

$$p_{i,t} - r_{i,1,t} \ge \underline{P}_{i,t} x_{i,t}, \forall i, \forall t, \tag{10}$$

$$-\left(RR_i + \overline{P}_{i,t} w_{i,t}\right) \le p_{i,t} - p_{i,t-1} \le RR_i + \left(\underline{P}_{i,t} - \frac{RR_i}{2}\right) u_{i,t}, \forall i, \forall t, \tag{11}$$

$$0 \le r_{i,z,t} \le \overline{R}_{i,z,t}, \forall i, \forall z, \forall t. \tag{12}$$

In the above, the maximum and minimum power output for unit i at time t are denoted as $\overline{P}_{i,t}$ and $\underline{P}_{i,t}$, respectively. The ramp rate for unit i is denoted as RR_i. For a certain unit, the energy generated within 24 hours (or 96 periods) is limited by its maximum daily energy available TE_i:

$$\sum_{t=1}^{96} \frac{p_{i,t}}{4} \le TE_i. \tag{13}$$

Similarly, the daily start-up times are limited by TS_i:

$$\sum_{t=1}^{96} u_{i,t} \le TS_i. \tag{14}$$

Virtual transactions include virtual generations and virtual demands. They are subject to capacity constraints:

$$0 \le m_{v,t} \le \overline{M}_{v,t}, v \in VG, \forall t, \tag{15}$$

$$\underline{M}_{v,t} \le m_{v,t} \le 0, v \in VD, \forall t, \tag{16}$$

where VG and VD are the sets of virtual generations and virtual demands, respectively. The generation level (or demand) of virtual transaction v at time t is denoted by $m_{v,t}$, and is limited by $\overline{M}_{v,t}$ (or $\underline{M}_{v,t}$). Similarly, dispatchable demand bid y has a maximum limit $\overline{D}_{y,t}$ on its level $d_{y,t}$ at period t.

$$0 \le d_{y,t} \le \overline{D}_{y,t}, y \in [1,Y], \forall t. \tag{17}$$

Virtual transactions and dispatchable demand bids are related to continuous variables only and have linear costs $C_{v,t}^V m_{v,t}$ and $C_{y,t}^Y d_{y,t}$, respectively.

Objective Function

The objective function is formulated as:

$$\min_{\substack{\delta,u,p,\\ r,x,y}} \left\{ \begin{aligned} &\sum_{i=1}^{I}\sum_{t=1}^{T}\left[\sum_{s=1}^{S} C_{i,s,t}^{Start} b_{i,s,t} + C_{i,t}^{NL} x_{i,t} + C^E(p_{i,t}) \right. \\ &\left. + \sum_{z=1}^{Z} C_{i,z,t}^{R} r_{i,z,t}\right] + \sum_{v=1}^{V}\sum_{t=1}^{T} C_{v,t}^{V} m_{v,t} - \sum_{y=1}^{Y}\sum_{t=1}^{T} C_{y,t}^{Y} d_{y,t} \\ &+ c^T \sum_{t=1}^{T}\sum_{l=1}^{L}\left(\overline{s}_{t,l} + \underline{s}_{t,l}\right) \end{aligned} \right\}. \tag{18}$$

In the above, three types of start-ups (hot, intermediate and cold) indexed by s ($1 \le s \le S$) are represented by the binary variable $b_{i,s,t}$, which is selected according to the values of $u_{i,t}$ and $w_{i,t}$ based on Equations (2)-(3) of [22]. Costs and penalties within (18) include costs from conventional units (start-up costs $\{C_{i,s,t}^{Start} b_{i,s,t}\}$, no-load costs $\{C_{i,t}^{NL} x_{i,t}\}$, piece-wise linear generation costs $\{C^E(p_{i,t})\}$, and reserve costs $\{C_{i,z,t}^R r_{i,z,t}\}$); costs from virtual transactions $\{C_{v,t}^V m_{v,t}\}$ and dispatchable demand bids $\{C_{y,t}^Y d_{y,t}\}$; and linear soft transmission capacity penalties. The problem is subject to system demand constraints (1), system reserve constraints (2), transmission capacity constraints (3)-(5), and all unit-level constraints (6)-(17). The overall problem is an MILP problem since the objective function and all constraints are linear, and both binary and continuous variables are included.

IV. Solution Methodology

Subsection IV.A presents the key steps of decomposing the problem into subproblems based on SAVLR with a few major modifications. In subsection IV.B, ordinal optimization concepts are introduced to provide "good enough" feasible subproblem solutions so as to avoid solving subproblems as MILP problems. In subsection IV.C, coordination of subproblem solutions, algorithm initialization, and finding feasible solutions are presented. Subsection IV.D presents a parallel version of the method to further reduce the CPU time.

A. Problem decomposition

This subsection presents the decomposition process based on SAVLR. The system-wide constraints are firstly relaxed or softened, and then the relaxed problem is decomposed into subproblems by properly grouping conventional units, virtual transactions and dispatchable demand bids.

Relaxing or softening system-wide constraints

Unlike the approach presented in [14], not all system-wide constraints are relaxed here. Instead, only system demand constraints are relaxed by using the Lagrangian multipliers $\lambda = (\lambda_1, \ldots, \lambda_T)'$, where each element λ_t is a scalar; and their violations are penalized with the adjustable penalty coefficient c^D. To avoid having an excessive number of multipliers, soft transmission capacity constraints are not relaxed, but are allowed to be violated with a fixed penalty coefficient c^T following the approach of [15]. As for system reserve constraints, although they are modeled as hard constraints in (2), they are also treated as soft during the iterative multiplier updating process and are not relaxed. Specifically, the non-negative slack variable $s_{z,t}^R$ is introduced, and the original system reserve constraints (2) are softened as:

$$\sum_{i=1}^{I} r_{i,z,t} + s_{z,t}^R \ge R_{z,t}, 1 \le z \le Z, \forall t. \quad (19)$$

Here, an adjustable penalty coefficient c^R is used to penalize the positive value of slack variable $s_{z,t}^R$, slightly different from the approach of [15]. By dynamically increasing c^R when the original system reserve constraints are violated, feasibility can be emphasized. At the final stage of the solution process to find feasible solutions, the original system reserve constraints (2) are required to be satisfied.

With the above, the relaxed problem is:

$$\min_{\substack{\lambda,\delta,u,p,\\ r,x,y,s}} \left\{ \begin{array}{l} \sum_{i=1}^{I}\sum_{t=1}^{T}\left[\sum_{s=1}^{S} C_{i,s,t}^{Start} b_{i,s,t} + C_{i,t}^{NL} x_{i,t} + C^E(p_{i,t}) \right. \\ \left. +\sum_{z=1}^{Z} C_{i,z,t}^R r_{i,z,t}\right] + \sum_{v=1}^{V}\sum_{t=1}^{T} C_{v,t}^V m_{v,t} - \sum_{y=1}^{Y}\sum_{t=1}^{T} C_{y,t}^Y d_{y,t} \\ +\sum_{t=1}^{T}\lambda_t g_t(p,m,d) + c^D \sum_{t=1}^{T} | g_t(p,m,d) | \\ +c^R \sum_{z=1}^{Z}\sum_{t=1}^{T} s_{z,t}^R + c^T \sum_{t=1}^{T}\sum_{l=1}^{L}\left(\overline{s}_{t,l} + \underline{s}_{t,l}\right) \end{array} \right\}, \quad (20)$$

where

$$g_t(p,m,d) \equiv \sum_{n=1}^{N} D_{n,t} - \sum_{i=1}^{I} p_{i,t} - \sum_{v=1}^{V} m_{v,t} + \sum_{y=1}^{Y} d_{y,t}, \forall t, \quad (21)$$

is the violation of demand constraints, and is penalized with the coefficient c^D. The relaxed problem is subject to softened system reserve constraints (19), transmission capacity constraints (3)-(5), and unit-level constraints (6)-(17).

<u>*Formulating subproblems*</u>

Following [14], conventional units in the relaxed problem (20) are divided into J subproblems based on areas (a subproblem j is formed by collecting all terms in (20) related to area j $(1 \le j \le J)$). Virtual transactions and dispatchable demand bids can also be divided into these subproblems based on areas. This, however, will cause subproblem solutions to drastically change across iterations because virtual transactions and dispatchable demand bids do not have discrete decision variables and are only subject to simple bounds (15)-(17). Consequently, their solutions are sensitive to the values of Lagrangian multipliers. This, in turn, may cause significant changes of multipliers across iterations, resulting in slow convergence. Therefore, different from conventional units, all virtual transactions and dispatchable demand bids are included in every subproblem. The objective function of a subproblem is formed by collecting all the terms in (18) associated with decision variables belonging to that subproblem while fixing decision variables of conventional units belonging to other subproblems at their latest available values. For compactness of expression, subscripts "*j*" and "*-j*" are used to indicate whether variables belong to subproblem *j* or not. For example, I_j is the set of conventional units belonging to subproblem *j*, and I_{-j} is the set of conventional units not belonging to subproblem *j*. The objective function of subproblem j at iteration k is thus as follows:

$$\min_{\substack{\lambda,\delta,u,p,\\ r,x,y,s}} \left\{ \begin{array}{l} \sum_{i\in I_j}\sum_{t=1}^{T}\left[\sum_{s=1}^{S} C_{i,s,t}^{Start} b_{i,s,t}^k + C_{i,t}^{NL} x_{i,t}^k + C^E(p_{i,t}^k) \right. \\ \left. +\sum_{z=1}^{Z} C_{i,z,t}^R r_{i,z,t}^k\right] + \sum_{v=1}^{V}\sum_{t=1}^{T} C_{v,t}^V m_{v,t}^k - \sum_{y=1}^{Y}\sum_{t=1}^{T} C_{y,t}^Y d_{y,t}^k \\ +\sum_{t=1}^{T}\lambda_t^k g_t(p_j^k, p_{-j}^{k-1}, m^k, d^k) + c^{D,k} \sum_{t=1}^{T} (q_t^{D,+} + q_t^{D,-}) \\ +c^R \sum_{z=1}^{Z}\sum_{t=1}^{T} s_{z,t}^{R,k} + c^T \sum_{t=1}^{T}\sum_{l=1}^{L}\left(\overline{s}_{t,l}^k + \underline{s}_{t,l}^k\right) \end{array} \right\}. \quad (22)$$

In the above,

$$g_t(p_j^k, p_{-j}^{k-1}, m^k, d^k) \equiv \sum_{n=1}^{N} D_{n,t} - \sum_{i\in I_j} p_{i,t}^k - \sum_{i\in I_{-j}} p_{i,t}^{k-1} - \sum_{v=1}^{V} m_{v,t}^k + \sum_{y=1}^{Y} d_{y,t}^k, \forall t, \quad (23)$$

indicates the violation of demand constraints. The associated absolute-value penalty term is linearized with the introduction of the non-negative continuous variables $q_t^{D,+}$ and $q_t^{D,-}$ and the following constraint as explained on the page 63 of [25]:

$$q_t^{D,+} - q_t^{D,-} = g_t(p_j^k, p_{-j}^{k-1}, m^k, d^k), \forall t. \quad (24)$$

In the above, $q_t^{D,+}$ and $q_t^{D,-}$ represent the violation of demand constraint at the positive and negative side, respectively. For system reserve and transmission capacity constraints, the decision variables of units belonging to other subproblems are fixed at their latest available values, and constraints (19) and (3)-(5) are rewritten as:

$$\sum_{i\in I_j} r_{i,z,t}^k + \sum_{i\in I_{-j}} r_{i,z,t}^{k-1} + s_{z,t}^{R,k} \ge R_{z,t}, \forall z, \forall t. \quad (25)$$

$$\tilde{f}_{t,l}^k - \overline{s}_{t,l}^k \le \overline{F}_l \;, \forall t, \forall l, \quad (26)$$

$$\tilde{f}_{t,l}^k + \underline{s}_{t,l}^k \ge \underline{F}_l \;, \forall t, \forall l, \quad (27)$$

$$\tilde{f}_{t,l}^k = \sum_{n=1}^{N} \alpha_{n,l}\left(\sum_{i\in I_j} p_{i,t}^k + \sum_{i\in I_{-j}} p_{i,t}^{k-1} + \sum_{v\in V_n} m_{v,t}^k - \sum_{y\in Y_n} d_{y,t}^k - D_{n,t}^k\right), \forall t, \forall l, \quad (28)$$

This subproblem is subject to the updated system reserve constraints (25), transmission capacity constraints (26)-(28), and unit-level constraints (6)-(17). It is still an MILP problem.

B. Quick searching process for good-enough feasible subproblem solutions

In SAVLR, subproblems are solved to satisfy the following surrogate optimality condition (Equation (14) of [14]):

$$\begin{array}{l} \tilde{L}(\lambda^k, b_j^k, x_j^k, p_j^k, p_{-j}^{k-1}, m^k, d^k) \\ \quad < \tilde{L}(\lambda^k, b_j^{k-1}, x_j^{k-1}, p_j^{k-1}, p_{-j}^{k-1}, m^{k-1}, d^{k-1}), \end{array} \quad (29)$$

where $\tilde{L}(\lambda^k, b_j^k, x_j^k, p_j^k, p_{-j}^{k-1}, m^k, d^k)$ is the "surrogate dual value" at iteration *k*, and is given by:

$$\begin{array}{l} \tilde{L}(\lambda^k, b_j^k, x_j^k, p_j^k, p_{-j}^{k-1}, m^k, d^k) \equiv \\ \sum_{i\in I_j}\sum_{t=1}^{T}\left[\sum_{s=1}^{S} C_{i,s,t}^{Start} b_{i,s,t}^k + C_{i,t}^{NL} x_{i,t}^k + C^E(p_{i,t}^k) + \sum_{z=1}^{Z} C_{i,z,t}^R r_{i,z,t}^k\right] \\ +\sum_{v=1}^{V}\sum_{t=1}^{T} C_{v,t}^V m_{v,t} - \sum_{y=1}^{Y}\sum_{t=1}^{T} C_{y,t}^Y d_{y,t} + \sum_{t=1}^{T}\lambda_t^k g_t(p_j^k, p_{-j}^{k-1}, m^k, d^k) \\ +c^{D,k}\sum_{t=1}^{T}(q_t^{D,k,+} + q_t^{D,k,-}) + c^R \sum_{z=1}^{Z}\sum_{t=1}^{T} s_{z,t}^{R,k} + c^T \sum_{t=1}^{T}\sum_{l=1}^{L}\left(\overline{s}_{t,l}^k + \underline{s}_{t,l}^k\right). \end{array} \quad (30)$$

The right-hand side of (29) is similarly defined.

To satisfy (29), subproblems are normally solved by using B&C. This is generally acceptable from the computational standpoint, since subproblems are much smaller than the original problem. However, for large sub-hourly subproblems, e.g., MISO's, B&C may suffer from poor performance. This difficulty is resolved by a novel exploitation of the Ordinal Optimization concepts.

Inspired by the OO concepts introduced in subsection II.D, a novel idea to significantly speed up the subproblem solving process is as follows. Rather than solving a subproblem by using an MILP method such as B&C, "good-enough" feasible solutions that satisfy the surrogate optimality condition (29) can be quickly obtained through "ordering" solution candidates. These candidates can be derived, for example, by modifying solutions obtained in the previous iterations for feasibility using heuristics, e.g., neighborhood search [19]. They can also be obtained by solving a crude subproblem, e.g., an LP relaxed version, and then making solutions feasible to the subproblem using heuristics. Solution candidates are arranged base on the ascending order of the associated surrogate dual values. A good enough subproblem solution is then obtained if the solution candidate with smallest surrogate dual value satisfies (29) the surrogate optimality condition. These ways to obtain subproblem solutions are much more computationally efficient than by using B&C. Only when good-enough solutions cannot be obtained, B&C is used. This approach is therefore much faster than solving subproblems exclusively by using B&C as will be demonstrated in Section V.

C. Coordination of subproblem solutions, initialization and finding feasible solutions

This subsection presents the coordination of subproblem solutions through updating multipliers and penalty coefficients; initialization of subproblem solutions, multipliers and penalty coefficients; and finding feasible solutions at the termination of iterative subproblem solving and multiplier updating process.

<u>Updating multipliers and penalty coefficients</u>

If the surrogate optimality condition (29) is satisfied by the solution obtained from the OO concepts, the surrogate subgradient is obtained as the values of $g_t(p_j^k, p_{-j}^{k-1}, m^k, d^k)$ in (23), and multipliers λ are updated following (17) of [14]:

$$\lambda_t^{k+1} = \lambda_t^k + s^k g_t(p_j^k, p_{-j}^{k-1}, m^k, d^k), \forall t. \tag{31}$$

In (31), the step size s^k is obtained following (18-19) of [14]. The penalty coefficient c^D is updated based on (20) of [14]:

$$c^{D,k+1} = \beta c^{D,k}, \ \beta > 1. \tag{32}$$

When (29) is not satisfied by the solution obtained from the OO concepts, B&C is used to solve the problem. If (29) is satisfied by the B&C solution, the multipliers λ and penalty coefficient c^D are updated by (31)-(32). Otherwise, the above updating process is skipped, and the next subproblem is solved. However, if (29) cannot be satisfied for all the J subproblems within a major iteration (i.e., all subproblems are solved once), then the penalty coefficient c^D is deemed to be too large, and is reduced by following (21) of [14].

$$c^{D,k+1} = \frac{c^{D,k}}{\beta}, \ \beta > 1. \tag{33}$$

As mentioned in subsection IV.A, the penalty coefficient on transmission capacity constraints c^T is a fixed value, and the penalty coefficient on system reserve constraints c^R is dynamically increased to minimize the violation of original reserve constraints. If any slack variable $s_{z,t}^R$ is positive, c^R is increased by multiplying a constant α (>1); and remains the same otherwise, i.e.,

$$c^R = \begin{cases} c^R \times \alpha & \text{if } s_{z,t}^R \neq 0, \ \alpha > 1 \\ c^R & \text{otherwise.} \end{cases} \tag{34}$$

The above process will lead to the convergence of multipliers λ to the optimal λ^* as presented in Theorem 1 of [14].

<u>Initializing subproblems solutions, multipliers and penalty coefficients</u>

The initialization of SAVLR parameters is also implemented by using the "good enough" concept. Before the iterative subproblem solving process, the hourly LP-relaxed UC problem is solved. Its solution is rounded and duplicated to all 15-minute intervals within the same hour as the initial subproblem solutions. They are modified to provide solution candidates for the first major iteration as presented in subsection IV.B. Lagrangian multipliers are initialized by using the results obtained from the hourly LP relaxed UC problem as well. The initial penalty coefficients are set to be an order of magnitude higher than multiplier values.

<u>Finding feasible solutions</u>

The iterative subproblem solving and multiplier updating process terminates when stopping criteria are satisfied, e.g., the gap calculated against a lower bound is less than a certain percentage, the time limit is reached, or each subproblem has been solved for a certain number of times. With system-level constraints relaxed or softened, subproblem solutions, when put together, may not satisfy the original constraints (1)-(17). A feasible solution is then constructed by using heuristics. For example, subproblem solutions are adjusted by using neighborhood search (e.g., the one embedded in Gurobi or CPLEX); or a portion of the binary variables is fixed at subproblem solution values, and the remaining decision variables are solved by using B&C. To measure the quality of a feasible solution, the best known lower bound obtained by using B&C in advance is used to calculate the optimality gap

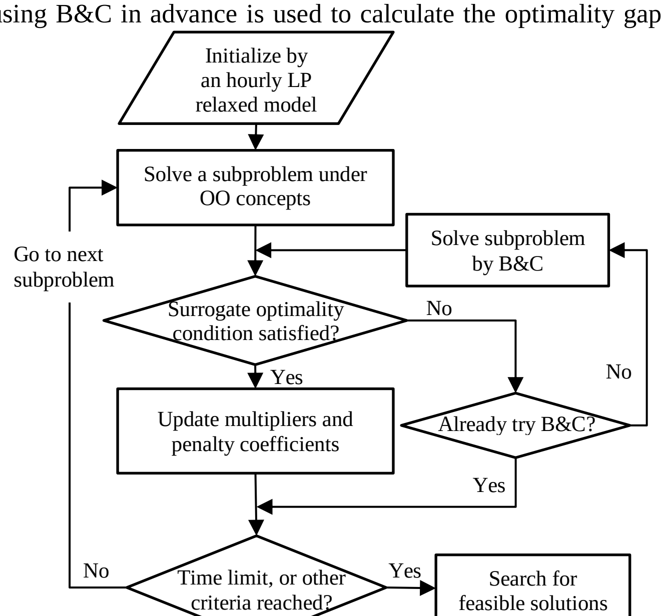

Fig. 1. Flowchart of the sequential SAVLR+OO+B&C

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2021.3137842, IEEE Transactions on Power Systems

following (17) in [15].

The above approach synergistically incorporates SAVLR, Ordinal Optimization, and B&C (SAVLR+OO+B&C), and the flow chart is presented in Figure 1. With the unique feature of SAVLR that subproblems are not required to be fully optimized, the OO concepts significantly speed up the subproblem solving process by not using the MILP methods unless needed. Furthermore, with the convergence condition (29) satisfied at most iterations, the quality of the feasible solution obtained at the end is generally good – similar to that of solutions obtained by using B&C to solve subproblems, even though the quality of subproblem solutions may not be as good as that obtained by using B&C. This will be demonstrated in numerical testing of Section V.

D.Parallelization of the method

To further reduce the CPU time, a parallel version of the approach is developed. The idea is to build subproblem models in parallel, and solve them in parallel at each iteration. Results from subproblems at each iteration are then merged to form a combined solution to update Lagrangian multipliers and penalty coefficients. There are, however, several difficulties. First, solving all subproblems in parallel at an iteration may lead to significant zigzagging of multipliers. This is precisely one of the major difficulties of the traditional LR: when all subproblems are solved, subgradient, rather than surrogate subgradient, are obtained. With "ridges" in the dual function, subgradient may change drastically across iterations, leading to multiplier zigzagging across ridges and slow convergence. Second, as explained in subsection IV.A, all virtual transactions and dispatchable demand bids are included in each subproblem. There are thus multiple values for each transaction or bid after solving multiple subproblems in parallel. Which one should be used? Finally, even if each subproblem solved in parallel satisfies the surrogate optimality condition, the merged solution might not, leading to convergence difficulties.

To overcome the above-mentioned difficulties, the following steps are taken. First, a small subset of subproblems (10% to 40% based on testing experience) is solved in parallel in a batch in a round-robin manner following the suggestion of [15]. To resolve the second and third difficulties identified above, a solution checking process is developed when merging subproblem results to form a combined solution. The results for conventional units from subproblems in the batch are combined in multiple ways. By fixing virtual transactions and dispatchable demand bids at the values obtained from the previous batch of subproblems, the merged solutions can be checked whether the surrogate optimality condition (29) is satisfied. If (29) is satisfied, values of virtual transactions and dispatchable demand bids are then determined by solving an extra LP problem with all units' variables fixed. If no combined solution satisfies (29), the solution of the subproblem with the lowest surrogate dual value (30) is selected, and there is no need to solve the extra LP problem. For example, suppose that three subproblems are solved in a batch. A combined solution obtained by merging three subproblem results is first checked to see if the surrogate optimality condition is satisfied. If so, the extra LP problem is solved, and multipliers and the penalty coefficient c^D are updated, and then the next batch of three subproblems is solved in parallel. If not, a combined solution

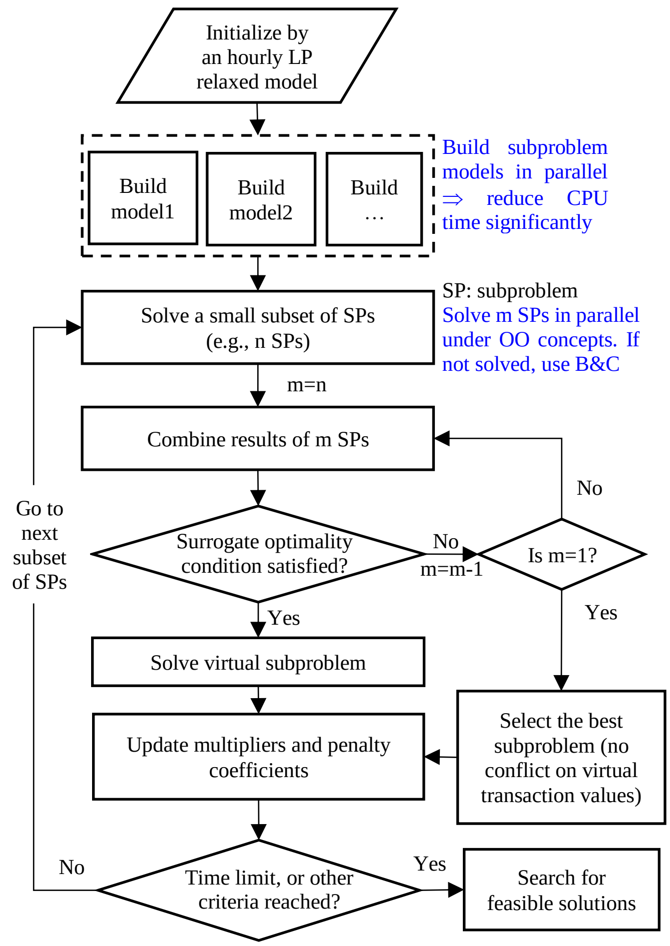

Fig. 2. Flowchart of the parallel SAVLR+OO+B&C

obtained by merging any two subproblem solutions is checked to see if the surrogate optimality condition is satisfied, and the process repeats. If no merged solution satisfies (29), then the solution of the subproblem with the lowest surrogate dual value is selected, and there is no need to solve the extra LP problem. The flow chart of the parallel version is presented in Figure 2.

V. Numerical Testing

Our method, both the sequential and the parallel versions, have been implemented by using Gurobi 7.5.0 and Python 2.7. Testing has been performed on the HIPPO platform of a MISO server with Intel Xeon @2.3GHz, 64GB RAM and 24 cores with Linux Redhat 6.6. Two examples of MISO's UC problems are considered with 15 minutes as the time interval over 36 hours. Example 1 is used to demonstrate the computational efficiency of our new method. In Example 2, three additional MISO cases with different numbers of units and locations of virtual transactions are tested to demonstrate the robustness of our method. For both examples, high quality solutions are difficult to obtain by using B&C alone within 20 minutes (1200s) or 30 minutes (1800s) even for hourly UC.

Example 1

In this example, a MISO 15-min interval UC problem is considered over 36 hours. There are 1,105 conventional units, 15,843 virtual transactions, 75 dispatchable demand bids, and 227 transmission lines. Following the process of Section IV, the problem is decomposed into 10 subproblems, each with roughly

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2021.3137842, IEEE Transactions on Power Systems

110 units and all virtual transactions and dispatchable demand bids. The initialization of our method is as follows: the average value of initial multipliers is \$12.56/MW; the initial penalty coefficient is \$125.6/MW; the fixed penalty coefficient is \$2000/MW; and the initial adjustable penalty coefficient for not meeting system reserve constraints is \$500/MW with the growth rate α equal to 1.01. The problem is solved by using B&C, SAVLR+B&C (solving subproblems sequentially by using B&C), and both sequential and parallel versions of the new method. In the parallel version, 3 subproblems are solved in parallel in a round robin manner. The stopping criterion is set as 1% of the gap (calculated by the feasible solution cost and the best known lower bound obtained by B&C in advance).

TABLE I
PERFORMANCE OF DIFFERENT METHODS FOR EXAMPLE 1

Methods	Solving Time (s)	CPU Time (s)	Gap (%)
B&C	5211	5443	0.90
SAVLR +B&C	2985	4086	0.90
SAVLR+ OO+B&C (sequential)	1484	3237	0.77
SAVLR+ OO+B&C (parallel)	979	1639	0.84

The overall results are summarized in Table I. As can be seen from the table, B&C obtains a feasible solution with a gap of 0.90% after more than 5,000s. For SAVLR+B&C, a feasible solution with a gap of 0.90% is obtained after 4,000s. For the sequential version of our approach, as shown in the third row of Table I, a feasible solution with a gap of 0.77% is obtained after 3,237s. The total solving time is 1,484s, and the rest are model loading and miscellaneous times. For the parallel version of our approach, as shown in the last row, a near-optimal solution with a gap of 0.84% is obtained after 1,639s.

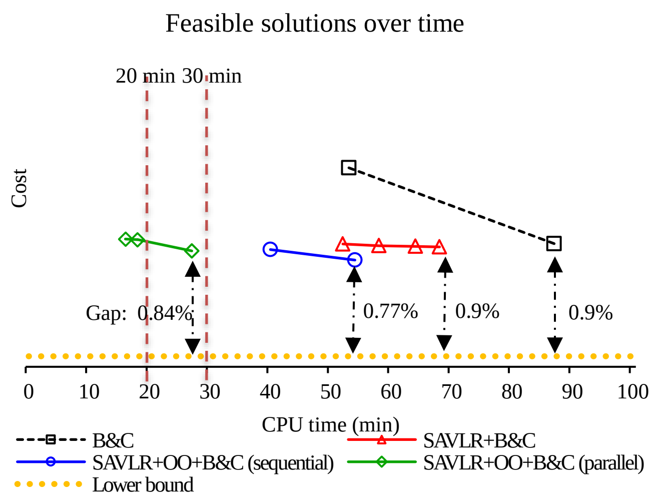

Fig. 3. Comparison of the feasible solutions obtained by SAVLR+OO+B&C, Pure B&C and SAVLR+B&C over time.

The feasible solutions obtained by using different methods over time are compared in Figure 3. Only the parallel version of our approach satisfies the stopping criterion of 1% gap within the required 1800s (i.e., 30 minutes). In the testing, the new method (both sequential and parallel versions) obtains good feasible solutions after solving each subproblem only twice (i.e., after two "major iterations"), same as that of the SAVLR+B&C. Within the new method, B&C was never called to solve subproblems. Rather, good-enough feasible subproblem solutions are always obtained by modifying existing solutions obtained from previous iterations (as part of the Gurobi "presolving" process). By doing this, the average time to obtain a good-enough feasible solution (both sequential and parallel versions) is 53s, which is much less than solving a subproblem by using B&C of 162s. Moreover, by applying parallelization, the overhead of model building and miscellaneous time is much reduced from 1,753s to 660s, and the total solving time is reduced from 1,484s to 979s. These results show that the OO concepts significantly speed up the subproblem solving process. Furthermore, even though our subproblems are not solved by using B&C, both sequential and parallel versions obtain high quality overall solutions (within 1% of the gap) after the same number of major iterations as that of SAVLR+B&C. Our new method thus significantly outperforms B&C and SAVLR+B&C.

Example 2

To demonstrate the robustness of our method, three additional cases roughly of the size of Example 1 but with different days of the MISO system are tested. Characteristics of test cases are summarized in Table II.

TABLE II
CHARACTERISTICS OF CASE 1, 2 AND 3

	# of units	# of virtual transactions	# of transmission constraints each interval
Case 1	1,109	16,504	220
Case 2	1,118	14,955	226
Case 3	1,102	14,482	235

These three cases are solved by using B&C and the parallel version of our new method. Similar to that of Example 1, the problem is decomposed into 10 subproblems; 3 subproblems are solved in parallel. The initial values of multipliers and penalty coefficients are close to those values of Example 1. With 2 major iterations as the stopping criterion for our method, and 3,600s (1 hour) as the time limit for B&C, the testing results are summarized in Table III.

TABLE III
PERFORMANCE OF B&C AND OUR APPROACH FOR CASE 1, 2 AND 3

	Methods	Solving Time (s)	CPU Time (s)	Gap (%)
Case 1	B&C	2548	3600	2.00
	Our approach (parallel)	990	1409	1.10
Case 2	B&C	2787	3600	4.31
	Our approach (parallel)	638	993	3.09
Case 3	B&C	3089	3600	76.00
	Our approach (parallel)	619	1016	1.60

As can be seen from Table III, after 3,600s, B&C obtains a feasible solution with a gap of 2% for Case 1; a feasible solution with a gap of 4.31% for Case 2; and a feasible solution with a gap of 76% for Case 3. By using our parallel version, a feasible solution with a gap of 1.10% is obtained after 1,409s for Case 1; with a gap of 3.09% after 993s for Case 2; and with a gap of 1.60% after 1,016s for Case 3. Similar to that of Example 1, B&C was never called to solve subproblems in Cases 2 and 3. For Case 1, B&C was used only twice at the beginning of iterations. The above results thus demonstrate that our new approach obtains near-optimal solutions in a computationally efficient manner for different sub-hourly UC cases, and significantly outperforms B&C.

VI. Conclusion

This paper presents a novel decomposition and coordination approach. Instead of formally solving subproblems by using MILP methods, good-enough feasible subproblem solutions are obtained by modifying existing subproblem solutions or solving crude models based on the OO concepts. The approach leads to a significant reduction of computational requirements to obtain near-optimal solutions of a similar quality as compared to SAVLR+B&C.

Our new approach can be extended to solve stochastic sub-hourly UC with uncertainties upon further development. In our previous works, uncertainties were explicitly modeled as discrete Markov processes. Without considering transmission capacity constraints, stochastic hourly UC was solved by using B&C [26]. With transmission capacity constraints, a hybrid Markovian and interval approach was developed, and after linearization, the problem was again solved by using B&C [27]. Branch-and-Cut, however, is not able to solve large deterministic sub-hourly UC as evident from numerical testing results presented in Section V, not to mention stochastic sub-hourly UC. Our approach presented here is conceivable to solve stochastic sub-hourly UC with the OO concepts further extended to handle the complicated Markov processes. This belief is built on the fact that with decomposition and coordination, subproblem complexity is much reduced as compared to that of the original problem. Then with the OO concepts further extended to appropriately approximate the complicated Markov processes, subproblem solving can be fast.

Our method represents a new optimization concept, and will have vital implications on solving other complicated MILP problems in power systems and beyond. Our next work will be on stochastic sub-hourly UC.

VII. Acknowledgment

The authors would like to express their sincere appreciation to Professor Yu-Chi Ho of Harvard Univ. for his inspiration on combining mathematical optimization with Ordinal Optimization; and to Mr. Anbang Liu of Tsinghua Univ. in Beijing for his work on combining mathematical optimization with OO to solve generalized assignment problems.

References

[1] J. Cochran, M. Miller, O. Zinaman, M. Milligan, D. Arent, B. Palmintier, M. O'Malley, S. Mueller, E. Lannoye, A. Tuohy, B. Kujala, M. Sommer, H. Holttinen, J. Kiviluoma, and S. K. Soonee, "Flexibility in 21st Century Power Systems." *21st Century Power Partnership*, May. 2014.

[2] "FERC Order 764," Tech. Rep. Accessed: Aug. 26, 2021. [Online]. Available at: https://www.ferc.gov/sites/default/files/2020-06/OrderNo.764.pdf.

[3] S. Pineda, R. Fernández-Blanco, and J. Miguel Morales, "Time-adaptive unit commitment," *IEEE Transactions on Power Systems*, Vol. 34, No. 5, pp. 354–359, May. 2019.

[4] J. P. Deane, G. Drayton, and B. P. Ó. Gallachóir, "The impact of sub-hourly modeling in power systems with significant levels of renewable generation," *Applied Energy*, Vol. 113, pp. 152–158, Jan. 2014.

[5] J. Wang, J. Wang, C. Liu, and J. P. Ruiz, "Stochastic unit commitment with sub-hourly dispatch constraints," Applied Energy, vol. 105, pp. 418–422, 2013.

[6] M. Kazemi, P. Siano, D. Sarno, and A. Goudarzi, "Evaluating the impact of sub-hourly unit commitment method on spinning reserve in presence of intermittent generators," *Energy*, Vol. 113, pp. 338–354, Oct. 2016.

[7] T. Kërçi, J. Giraldo, F. Milano, "Analysis of the impact of sub-hourly unit commitment on power system dynamics," *International Journal of Electrical Power & Energy Systems*, Vol. 119, Jul. 2020.

[8] B. Yan, P. B. Luh, T. Zheng, D. Schiro, M. A. Bragin, F. Zhao, J. Zhao, and I. Lelic, "A Systematic Formulation Tightening Approach for Unit Commitment Problems," *IEEE Transactions on Power Systems*, Vol. 35, No. 1, pp. 782–794, Jan. 2020.

[9] Y. Chen, A. Casto, F. Wang, Q. Wang, X. Wang, and J. Wan, "Improving large scale day-ahead security constrained unit commitment performance," *IEEE Transactions on Power Systems*, Vol. 31, No. 6, pp. 4732-4743, Feb. 2016.

[10] J. E. Mitchell, "Branch-and-cut," Wiley encyclopedia of operations research and management science. John Wiley & Sons, Inc, 2010.

[11] S. Virmani, E. C. Adrian, K. Imhof, and S. Mukherjee, "Implementation of a Lagrangian relaxation based unit commitment problem," *IEEE Transactions on Power Systems*, Vol. 4, No. 4, pp. 1373-1380, Nov. 1989.

[12] X. Guan, P. B. Luh, H. Yan, and J. A. Amalfi, "An optimization-based method for unit commitment," *Int. J. Elec. Power Energy System*, Vol. 14, No. 1, pp. 9–17, Feb. 1992.

[13] M. A. Bragin, P. B. Luh, J. H. Yan, N. Yu, and G. A. Stern, "Convergence of the surrogate Lagrangian relaxation method," *J. Optimization Theory and Applications*, Vol. 164, No. 1, pp. 173-201, Apr. 2015.

[14] M. A. Bragin, P. B. Luh, B. Yan, and X. Sun, "A scalable solution methodology for mixed-integer linear programming problems arising in automation," *IEEE Transactions on Automation Science and Engineering*, Vol. 16, No. 2, pp. 531-541, Apr. 2019.

[15] N. Raghunathan, M. A. Bragin, B. Yan, P. B. Luh, K. Moslehi, X. Feng, Y. Yu, C. Yu, C. Tsai, "Exploiting Soft Constraints within Decomposition and Coordination Methods for Sub-hourly Unit Commitment," Submitted and available at TechRxiv. Accessed: Aug. 26, 2021. Available at: https://www.techrxiv.org/articles/preprint/Exploiting_Soft_Constraints_within_Decomposition_and_Coordination_Methods_for_Sub-hourly_Unit_Commitment/12950414.

[16] M. Deng and Y. C. Ho, "An ordinal optimization approach to optimal control problems," *Automatica*, Vol. 35, pp. 331-338, Feb. 1999.

[17] Y. C. Ho, Q. Zhao and Q. Jia, Ordinal Optimization: Soft optimization for hard problems, Springer US, 2007.

[18] J. Wu, P. B. Luh, Y. Chen, B. Yan, and M. A. Bragin, "A Decomposition and Coordination Approach for Large Sub-hourly Unit Commitment," *2020 IEEE Power & Energy Society General Meeting (PESGM)*, Montreal, QC, Aug. 2020.

[19] Y. Chen, F. Pan, J. Holzer, E. Rothberg, Y. Ma and A. Veeramany, "A High Performance Computing Based Market Economics Driven Neighborhood Search and Polishing Algorithm for Security Constrained Unit Commitment," *IEEE Transactions on Power Systems*, Vol. 36, No. 1, pp. 292-302, Jan. 2021.

[20] X. Sun, P. B. Luh, M. A. Bragin, Y. Chen, J. Wan, F. Wang, "A novel decomposition and coordination approach for large day-ahead unit commitment with combined cycle units," *IEEE Transactions on Power Systems*, Vol. 33, No. 5, pp. 5297-5308, Sept. 2018.

[21] A. B. Liu, P. B. Luh, M. A. Bragin and B. Yan, "Ordinal-Optimization Concept Enabled Decomposition and Coordination of Mixed-Integer Linear Programming Problems," *IEEE Robotics and Automation Letters*, Vol. 5, No. 4, pp. 5051-5058, Oct. 2020.

[22] G. Morales-España, J. M. Latorre and A. Ramos, "Tight and compact MILP formulation for the thermal unit commitment problem," *IEEE Transactions on Power Systems*, Vol. 28, No. 4, pp. 4897-4908, Nov. 2013.

[23] NERC BAL-001-2. [Online]. Accessed: August. 26, 2021. Available at: https://www.nerc.com/files/BAL-001-2.pdf.

[24] Y. Chen, R. Leonard, M. Keyser, and J. Gardner, "Development of performance-based two-part regulating reserve compensation on MISO energy and ancillary service market," *IEEE Transactions on Power Systems*, Vol. 30, No. 1, pp. 142-155, May. 2014.

[25] J. Bisschop, AIMMS Optimization Modeling. Morrisville, NC, USA: Lulu Press, 2006.

[26] P. B. Luh, Y. Yu, B. Zhang, E. Litvinov, T. Zheng, F. Zhao, J. Zhao and C. Wang, "Grid Integration of Intermittent Wind Generation: a Markovian Approach," *IEEE Transactions on Smart Grid*, Vol. 5, No. 2, pp. 732-741, Mar. 2014, DOI 10.1109/TSG.2013.2268462.

[27] Y. Yu, P. B. Luh, E. Litvinov, T. Zheng, J. Zhao and F. Zhao, "Grid Integration of Distributed Wind Generation: Hybrid Markovian and Interval Unit Commitment," *IEEE Transactions on Smart Grid*, Vol. 6, No. 6, pp. 3061-3072, Nov. 2015, DOI: 10.1109/TSG.2015.2430851.