Bender's Decomposition for Optimization Design Problems in Communication Networks

Ahmed Ibrahim, Octavia A. Dobre, Telex M. N. Ngatched, and Ana García Armada

Abstract—**Various types of communication networks are constantly emerging to improve the connectivity services and facilitate the interconnection of various types of devices. This involves the development of several technologies, such as device-to-device communications, wireless sensor networks and vehicular communications. The various services provided have heterogeneous requirements on the quality metrics such as throughput, end-to-end latency and jitter. Furthermore, different network technologies have inherently heterogeneous restrictions on resources, e.g., power, interference management requirements, computational capabilities, etc. As a result, different network operations such as spectrum management, routing, power control and offloading need to be performed differently. Mathematical optimization techniques have always been at the heart of such design problems to formulate and propose computationally efficient solution algorithms. One of the existing powerful techniques of mathematical optimization is Benders decomposition (BD), which is the focus of this article. Here, we briefly review different BD variants that have been applied in various existing network types and different design problems. These main variants are the classical, the combinatorial, the multi-stage, and the generalized BD. We discuss compelling BD applications for various network types including heterogeneous cellular networks, infrastructure wired wide area networks, smart grids, wireless sensor networks, and wireless local area networks. Mainly, our goal is to assist the readers in refining the motivation, problem formulation, and methodology of this powerful optimization technique in the context of future networks. We also discuss the BD challenges and the prospective ways these can be addressed when applied to communication networks' design problems.**

I. Introduction

Nowadays, communication networks have become highly complex due to the requirement to support heterogeneous applications. This impacts the decision-making encountered in various network operations such as routing, resource allocation, and user association. Specific design algorithms, which depend on the network type, its characteristics, and the user/application quality-of-service requirements are hence needed.

Mathematical programming techniques are widely used decision-making tools regardless of the network type. A decision is made such that a certain criterion is optimized (e.g., end-to-end delay) given some system related constraints (e.g., link capacity restrictions). As the network size increases, the number of decision variables and constraints also increases

A. Ibrahim, O. A. Dobre and T. M. N. Ngatched are with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL, Canada, e-mails: (amaibrahim@mun.ca, odobre@mun.ca ,tngatched@grenfell.mun.ca)

A. G. Armada is with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Avda. de la Universidad 30 (28911 Leganés) Spain e-mail:ana.garcia@uc3m.es

and the computational complexity of the decision-making algorithm becomes the bottleneck.

To design efficient algorithms for large problems, an exploitation of any mathematical structure in the problem's formulation is crucial. An effective approach for that is the use of *decomposition methods*. In this case, the variables and constraints are decomposed into several subsets with the constraints in one subset mainly depending on the variables in the same subset. The idea of decomposition methods is to temporarily fix the linking variables or drop the complicating constraints; hence, the blocks are decoupled and the problem decomposes into several smaller problems.

Benders Decomposition (BD) is one of the popular decomposition schemes [1], because it exploits the structure of the problem and decentralizes the overall computational burden. It has been successfully applied in a variety of network types, as illustrated in Figure 1. Since its introduction [2], several extensions have been proposed for the BD method. In this article, we present the basic concepts of the four main BD variants, as well as some of their various existing applications in communications networks. Moreover, we discuss the technical challenges with BD and possibilities to overcome them when applied to communication networks.

II. Basics of Benders Decomposition

The basic idea behind BD is to decompose the problem to be solved into two simpler problems, namely the master problem (MP) and the auxiliary problem (or sub-problem). The MP is a relaxed version of the original problem, containing only a subset of the original variables and the associated constraints. Its solution yields a lower bound on the objective function of the problem. The auxiliary problem is the original problem with the variables obtained in the MP fixed. Its solution yields an upper bound on the objective function of the problem and is used to generate cuts for the MP. The MP and auxiliary problem are solved iteratively, until the upper and lower bound are sufficiently close.

Besides the original BD, also known as *classical Benders decomposition* (CBD), other variants have been developed and successfully applied in different types of data networks problems. These variants include:

1) Combinatorial BD (ComBD) [3, 4];
2) Multistage *l-shaped decomposition* [5];
3) Generalized BD (GBD) [6].

Figure 2 illustrates, at a high level, an abstract flowchart of how the four BD variants work. Table I shows different types of communication networks for which a variant of BD has been used to solve a particular design problem, as well as the key features of each variant.
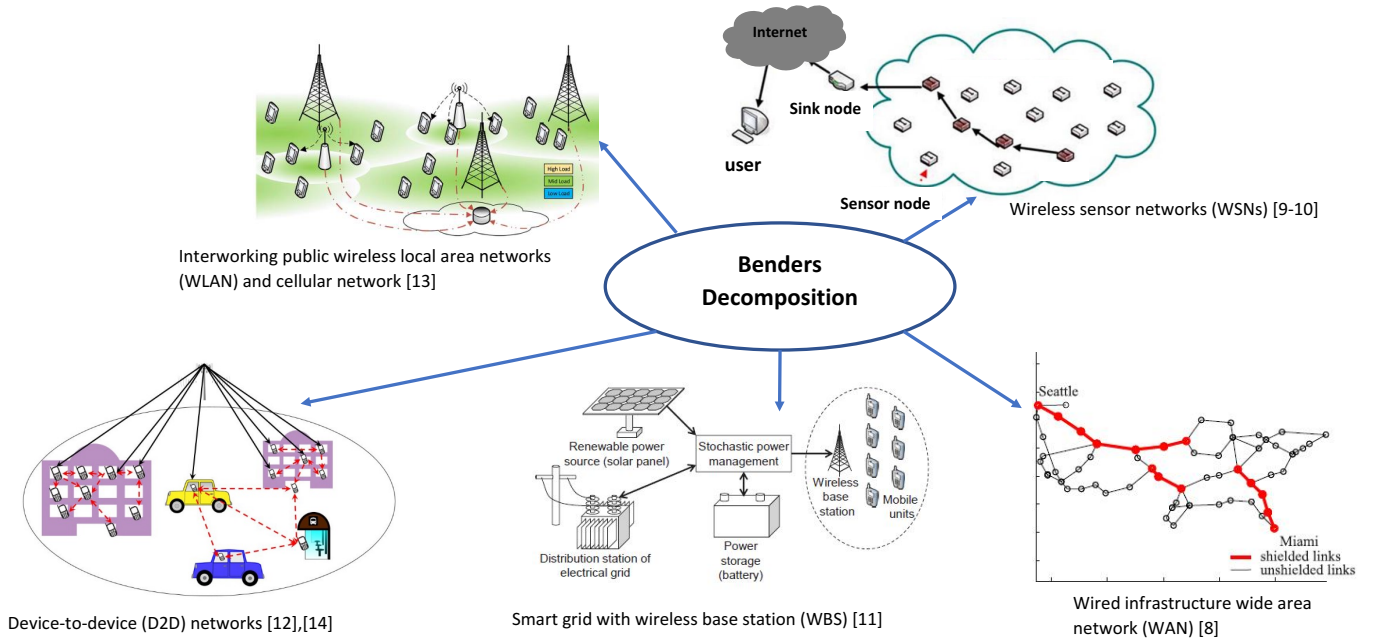
Figure 1: Different network types in which BD applications exist.

TABLE I: Different networks considered in this article to which BD is applied.

	CBD	ComBD	Multi-stage BD for stochastic programs	GBD
Key Features	i. Linear duality ii. Reformulation with the aid of extreme rays and extreme points	i. Deals specifically with integer sub-problems ii. Makes use of combinatorial cuts	i. BD applied more than once, recursively ii. Decision of a stage depends on that in the previous stage iii. Accounts for uncertainties	i. Deals with nonlinear programs specifically ii. Requires sub-problems to be convex iii. Exploits nonlinear duality iv. Two sub-problems are required for generating the two cut types
Applications	i. WSNs [9, 10] ii. Heterogeneous cellular wireless networks [7] iii. Wired infrastructure WANs [8]	WLANs [4]	Smart grid with WBS [11]	i. Interworking public WLAN and cellular networks [13] ii. D2D networks [12, 14]

A given problem can usually be modeled with different equivalent formulations. Formulations have a large impact on the BD performance, as they greatly affect the computational requirements. Formulations with strong *linear program* (LP) relaxations have better computational performance. The reasons are mainly because of the tighter root node, the lower number of fractional variables, and the stronger cuts generated. Tighter formulations can be obtained by adding additional constraints; however, this may result in a more time-consuming MP. Therefore, there is a trade-off between the reduction in the number of iterations and the additional difficulty of the sub-problem.

The BD method has strong ties to other types of decomposition, such as the Dantzig-Wolfe (DW) and Lagrangian optimization, when it comes to solving LPs. Solving an LP by DW is equivalent to solving its dual using BD. The BD is also equivalent to a cutting-plane method applied to the Lagrangian dual of the problem. On the other hand, in integer programming, the relationship between the decomposition methods is not obvious. A great advantage that BD has in this case is that it converges straight to the optimal of the *mixed integer linear program* (MILP) rather than to a relaxation of the problem, as DW decomposition and Lagrangian relaxation do. Therefore, BD needs not be embedded within a branch and bound framework.
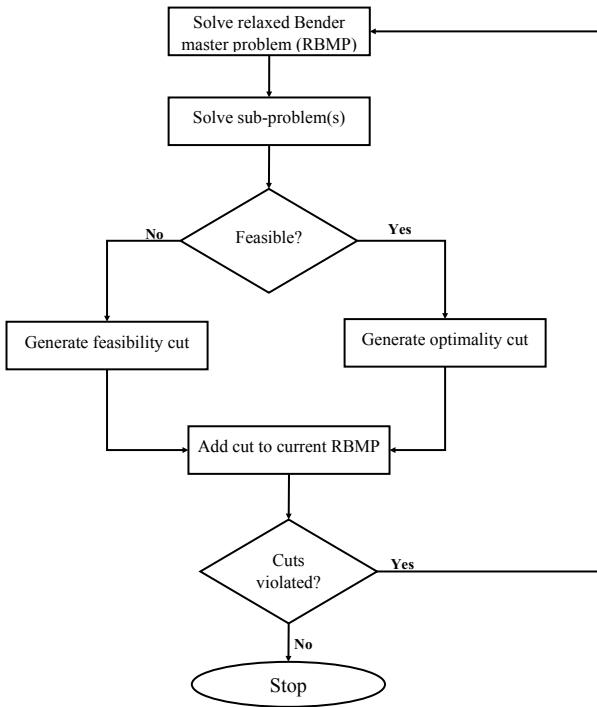
Figure 2: Flowchart representation of BD.

III. CLASSICAL BENDERS DECOMPOSITION (CBD)

CBD can be applied for an MILP that has a constraint set connecting both continuous and integer variables linearly. The *Benders master problem* (BMP) is obtained through the following steps:

1) Reformulating the problem into two nested minimization problems, where the outer minimization is over a discrete variable vector $\mathbf{x_1}$, while the inner minimization (IM) is only over a continuous variable vector $\mathbf{x_2}$.

2) Interchanging IM with its dual (IMD) using linear duality theory. The feasible region, F_{IMD}, of IMD is independent of the choice of $\mathbf{x_1}$, which means that if it is non-empty, IMD is either unbounded or feasible for any $\mathbf{x_1}$.

3)   i. An unbounded direction in F_{IMD} must be avoided, as it indicates an infeasibility of $\mathbf{x_1}$. To restrict movement in this direction, a *feasibility* cutting plane must be added.

  ii. If the inner IMD is feasible, the solution is an extreme point of F_{IMD}.

4) Another reformulation is done, where an auxiliary variable substitutes the linear IMD objective and a linear constraint set that is function in the vector of the extreme points of F_{IMD} is added. These constraints are the *optimality cuts*, which when added one after the other, tighten the lower bound on the objective function.

Since the complete enumeration of feasibility and optimality cuts is not possible, a relaxation and iterative approach is used to solve the BMP. In a given iteration, the main steps of the iterative procedure are as follows:

1) The *relaxed* BMP (RBMP) is solved for a subset of cuts to obtain a trial value $\mathbf{x_1} = \mathbf{x_1^*}$.

2) The IMD with $\mathbf{x_1^*}$ is solved. If the IMD is unbounded, a feasibility cut is generated. Otherwise, an optimality cut is generated.

3) If the cuts are violated by the current solution, they get added to the current RBMP and the process repeats.

CBD has been applied to different existing network design problems. These and their related CBD specifics, are summarized in Table II(a), and discussed next.

A. CBD for Base Station (BS) Association and Power Control

CBD is employed for user to BS association and power control in heterogeneous cellular networks in [7], with the aim of maximizing the overall system revenue measured by either the total data rate or the number of mobile stations in service. The signal-to-interference-and-noise-ratio (SINR) target on each link in service must be satisfied with the minimum total transmission power. It was shown that for certain weighting of the different objectives, the optimal solution of the resulting MILP maximizes the revenue and minimizes the power. The user to BS association is achieved by solving the RBMP iteratively. In the BS association and power control algorithm (BARN) [7], integer linear programming methods are used for solving the RBMP. In *accelerated* BARN [7], the integer requirements are relaxed and standard LP solution techniques are used. The feasibility pump heuristic [7] is used to convert the fractional to integer values. The power allocation is done by solving the sub-problem iteratively.

B. CBD for Critical Link Shielding

A problem that explores the shielding of critical links to enhance the robustness of a network, in which shielded links are resilient to failures, is investigated in [8]. The problem is to find the minimum cost shielding in order to resist both independent random failures and correlated failures, such that connectivity between all source-destination pairs is always maintained. The problem is modeled as a minimum cost network flow problem that takes into account different failure scenarios. The MP is composed of binary decision variables, where each indicates whether a link is shielded. The sub-problem is a zero objective optimization linear network flow problem that captures the possible failure patterns and guarantees a feasible flow for a corresponding shielding solution obtained from solving the linear integer RBMP.

Numerical Example: As shown/demonstrated in [8], applying modifications to the CBD can provide significant time enhancement. This is extremely useful because two failures may affect two disjoint sets of links, and the shielding decisions can be made in the same step to reduce the number of iterations. The number of violated constraints added before re-solving the master problem provides a tradeoff between the number of iterations and the running time of each iteration. Table II(b) shows the corresponding reproduced numerical results where the number of constraints is equal to the number of nodes in the network, which yields 50% reduction in the running time compared with the standard CBD.

Network	Design problem	Objective	Decision variables	Enhancements in CBD	RMBP	Sub-problem	Computational performance
• Heterogenous multicellular network [7]	• Joint user association and power control	• Throughput maximization and power consumption minimization in downlink	• Binary association decision variable • Continuous power decision variables	• Relaxation of integer requirements • Use of feasibility pump heuristic to restore feasibility	• User to BS association	• Power allocation LP	• **BS user association and power control (BARN)**: ○ 51 CBD Iterations • **Accelerated-BARN**: ○ 16 CBD iterations ○ 30% lower CPU time than BARN
• Wired infrastructure WAN [8]	• Shielding links to ensure connectivity of all nodes in the occurrence of failures	• Minimize the costs needed for shielding	• Binary selection variables indicating whether a link is shielded • Binary variables indicating whether a link is destroyed for a particular failure event • Continuous variables indicating the flow on a given link	• Multiple failures are checked instead of one	• Determines which links to shield	• A zero objective optimization linear network flow problem obtains a feasible flow for a corresponding shielding solution	• Modified BD is faster than the classical BD by 50 %
• WSNs [9]	• Optimal cluster head, sink nodes selection and flow routing in multi-sink data-gathering WSNs	• Maximize the lifetime of the network	• Binary variables for selecting cluster head nodes and sink nodes • Continuous variables for assigning routing flows	• Upper bound heuristic • Strengthened cuts • ϵ optimal approach	• Gives a tentative selection of cluster head and sink locations	• Determines flow routing and sensor energies	• Enhanced CBD outperforms CPLEX branch and cut significantly for all instance sizes, with about 90% on average in the largest instance experimented, at an optimality gap of 2%
• WSNs [10]	• Lifetime maximization problem in WSNs with a mobile sink	• Maximize the lifetime of the network	• Locations the sink should visit (binary) • Order with which the sink should visit a location (binary) • Flow volume on each link (continuous) • Time the sink must spend at a given location (continuous)	• N/A	• Determines the sink locations to be visited and their order	• Time duration the mobile sink should stay at a location • Flow assignment between two sensors	• Significant average CPU time savings as compared to direct solve by CPLEX

(a) Summary of design problems.

Number of nodes	CBD: Time (s)	Application-specific modification: Time (s)
10	2.26	1.54
15	20.89	10.85
20	90.77	34.96
25	270.69	103.32
20	684.83	195.2

(b) Running time comparison of CBD and its application-specific enhancement for infrastructure wired WAN link shielding [8].

TABLE II: Summary of design problems in which CBD is employed.

It is hence important to know that there are possibilities to enhance the convergence speed of BD in general, by exploiting application specific features of various communication networks. More on the BD challenges and potential ways to address those is discussed in Section VII.

C. CBD for Data Gathering in WSNs

Data gathering cluster-based WSNs are studied in [9], where a WSN with two sets of candidate sink nodes and cluster nodes is considered. The design problem is to find a selection from each set jointly with the network flow between the sensor node and the selected sinks, such that the lifetime of the network is maximized. The objective function represents the weighted sum of three terms: the average energy consumption, the range of remaining energy levels, and the fixed cost associated with locating the cluster heads. The MP is an integer linear program (ILP) which gives a tentative network configuration, i.e., the selection of a cluster head and sink locations. The sub-problem provides the optimal data routing and allocates the sensor energies for that fixed configuration. It is essentially a linear minimization problem that determines the data routing scheme from sensors to sinks via cluster heads and energy usage in the network. Three approaches are used in [9] to speed up the convergence of CBD: an upper bound heuristic algorithm;

strengthening the Benders cuts; and following an ϵ-optimal approach.

D. CBD for Routing and Mobile Sink Tour Optimization in WSNs

In [10], the problem of maximizing the lifetime of a WSN with a mobile sink is considered. The sink travels at a finite speed among a subset of possible sink locations to collect data from a stationary set of sensor nodes. The problem chooses a subset of sink locations for the sink to visit, finds a tour for the sink among the selected sink locations, and prescribes an optimal data routing scheme from the sensor nodes to each location visited. The sink's tour is constrained by the time it spends at each location collecting data. The CBD starts from a tour over a subset of the sink locations and reaches an optimal solution by adding appropriate Benders cuts. The sink's routing-travel decisions are made by the RBMP, which determines the locations to be visited and their order. The time duration the mobile sink has to stay at a particular location, as well as the flow between two sensors while the sink is at that location, are determined by the sub-problem.

IV. Combinatorial Benders Decomposition (ComBD)

When a subset of the projected variables must satisfy integrality requirements, the standard linear duality theory is no longer suitable for obtaining classical Benders cuts. A different framework is thus invoked for the cut generation in order to handle integer sub-problems. ComBD is suitable for a basic 0-1 ILP amended by a set of "conditional" linear constraints which come into effect if one of the 0-1 variables is 'on', i.e., $\mathbf{y}(j) = 1 \Rightarrow \mathbf{a}_i^T \mathbf{x} \geq b_i$ and possibly a set of "unconditional" constraints, where $\mathbf{x}$ is possibly a vector of continuous variables and $\mathbf{y}$ is a vector of binary variables. For such problems, the MP is a 0-1 integer program and the sub-problem is a feasibility optimization problem with no objective function. ComBD excludes the current MP solution from further consideration via *combinatorial cuts*, which usually take the form

$$\sum_{\forall j: \bar{\mathbf{y}}(j)=0} \mathbf{y}(j) + \sum_{\forall j: \bar{\mathbf{y}}(j)=1} (1 - \mathbf{y}(j)) \geq 1, \qquad (1)$$

and are used as feasiblity cuts in ComBD, where $\bar{\mathbf{y}}$ is the solution returned by the RBMP.

ComBD was used in [4] for user association and power control in WLANs. In this problem, each user equipment (UE) must be assigned to one access point (AP). Each AP can have different power levels (PLs) and can be powered off, but at most one PL must be chosen for each AP. The decisions are whether to use or not an AP, to assign a PL to each selected AP, and to assign exactly one AP to each UE. The objective is to minimize the aggregate power of all APs, while satisfying the nonlinear capacity constraints of each AP. The RBMP here solves for the power levels subject to the condition that only one power level per AP is assigned.

The sub-problem is a zero objective linear integer program. If this is infeasible, a cut as in equation (1) can be used as a feasibility cut. The results indicate that ComBD is often orders of magnitude faster than a benchmark cutting planes algorithm.

V. Multi-Stage Benders Decomposition for Stochastic Programs

This technique is suitable for solving *stochastic linear programming* problems, which can be modeled as multi-stage linear programs that consider uncertainty in at least some of the quantities involved in the problem [5]. Figure 3 shows a two-stage recourse problem versus a four stage deterministic problem.

The solution maximizes the expected value of the objective function over all possible scenarios. A multi-stage problem can be formulated as a single optimization problem known as the *deterministic equivalent*. A key approach for that is to discretize the possible future evolution of the state space. This yields a normal linear minimization problem, since the objective function is linear and the constraints are now finite.

A. Applying BD to Stochastic Programming Problems

The main idea is to split a very large LP problem, which is the deterministic equivalent of the stochastic programming problem, into smaller LP problems that are each much quicker to solve. The decomposition yields a single first stage problem and a number of second stage problems, one corresponding to each of the possible scenarios that could be realized in stage two.

Given an initial estimate of the first-stage variables, an attempt is made to solve each of the sub-problems assuming that this chosen value was the decision made in the first stage. There are two possibilities: either some of the sub-problems are infeasible or all sub-problems are feasible. For the former, feasibility cuts are added to the RBMP, while optimality cuts are added in the latter.

B. Stochastic Power Management for WBSs in Smart Grids

An application exists in power management of WBS in a smart grid [11]. The uncertainty is in spot power price of the electrical grid, the power availability in a renewable source, and the demands. The power price from the smart grid is time-varying and increases during the peak hour. The optimization is performed for $|\mathcal{T}|$ periods, where $\mathcal{T}$ is a set of time periods and $|.|$ is the set cardinality. A random variable, $\tilde{\omega}_t$, captures the possible spot power prices from electrical grid, the possible power generated from the renewable power source, and the possible number of connections in the WBS at period t. The objective function minimizes the expected cumulative cost over all periods. The period-dependent decision variables are the amount of power bought from an electrical grid, the amount of power stored in a battery, and the amount of excess power to be sold back to an electrical grid at period t. The cost at period t is defined with the period-dependent decision variables and composite random variables $\tilde{\omega}_t$ and $\tilde{\omega}_{t+1}$. The objective function is a linear function in the decision variables and represents an aggregation of the cost of buying power from the electricity grid, the cost of losing renewable power stored in the battery less the cost of selling power back to the electrical grid.

The problem formulation renders a stochastic program, which is transformed into a deterministic equivalent formulation that is an LP. The expectation in the stochastic program gets replaced by the weighted sum of probabilities of scenarios. The MP solves for the amount of power to be sold back to the power grid, for certain consumption levels of battery and grid power. The solution obtained from the MP is tested in the sub-problem. If a solution is optimal within the acceptable tolerance, the algorithm stops. Otherwise, the nearly-optimal values of primal and dual variables from the sub-problems are used to generate Benders cuts that can improve the solution, and another iteration is performed.

VI. Generalized Benders Decomposition (GBD)

GBD generalizes CBD to solve nonlinear problems for which the auxiliary sub-problem is a convex program [6]. As in CBD, there are two decision vectors $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$, where $\mathbf{y}$ is a vector of complicating variables in the sense that the problem is a much easier optimization problem in $\mathbf{x}$ when $\mathbf{y}$ is temporarily held fixed. If $\mathbf{y}$ consists of integer
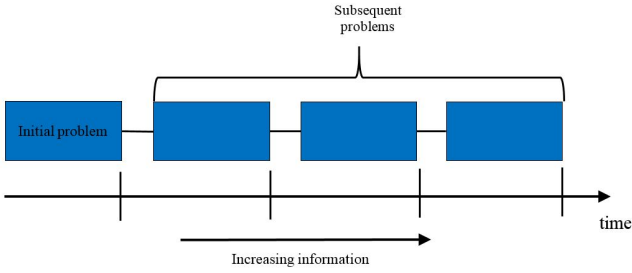
(a) Deterministic dynamic problem: There is no uncertainty about the evolution of information.
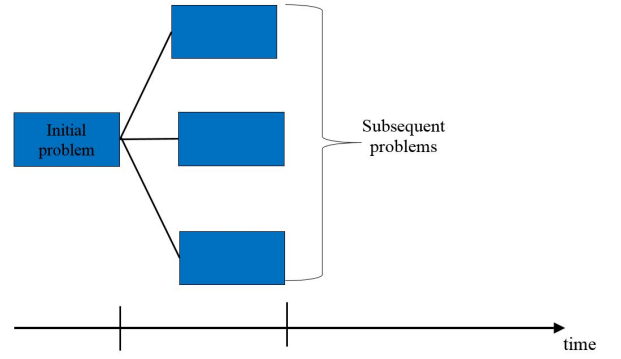
(b) Two-stage stochastic dynamic problem. Here there is uncertainty about which of the three possible successor states the problem moves into after the first decision, such that the decision must consider the uncertainty.

Figure 3: Multistage problems.

decision variables, then the problem is a *mixed integer non-linear program* (MINLP). GBD becomes largely effective if one or more of the following are fulfilled:

1) For a fixed $\mathbf{y}$, the problem divides into a number of independent optimization problems in a separate decision subvector of $\mathbf{x}$;
2) For a fixed $\mathbf{y}$, the problem assumes a well-known special structure, such as *shortest path tree problem* (STPT) or network flow;
3) Though the problem is not a convex program in $\mathbf{x}$ and $\mathbf{y}$ jointly, fixing $\mathbf{y}$ renders it so in $\mathbf{x}$.

Nonlinear convex duality is exploited to derive the families of cuts corresponding to those in CBD. In the GBD, to derive the RBMP, projection is invoked to project the problem on to $\mathbf{y}$, leading to $\min_{\mathbf{y}} \{v(\mathbf{y}) : \mathbf{y} \in \mathcal{Y} \cap \mathfrak{B}\}$, where $v(\mathbf{y})$ is a parametric function that corresponds to the optimum value of the problem under consideration for fixed $\mathbf{y}$, and $\mathfrak{B} \equiv \{\mathbf{y} : \mathbf{G}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}, \; for \; some \; \mathbf{x} \in \mathcal{X}\}$, where $\mathbf{G}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}$ is a constraint set in the original formulation. Non-linear duality is then invoked to derive the BMP, which has two sets of constraints that are both too large and involve an infimum term over $\mathbf{x} \in \mathcal{X}$. The natural strategy for solving the RBMP is ignoring all but a few of its many constraints, then iteratively adding the violated ones to the RBMP.

GBD has been applied in some network design problems. These and their related GBD specifics are summarized in Table III, and discussed in the sequel.

A. GBD for Joint Power Control and Routing in Multihop D2D Networks

In our work in [12], multihop relaying is considered to deliver the packets from the BS to the UEs whose direct link is in outage, i.e., whose signal-to-noise-ratio (SNR) is below an acceptable threshold. There is interference from the D2D relay transmissions on uplink cellular users, which must be managed. Furthermore, the resource blocks (RBs) for D2D sidelinks (SLs) are reused across the cell, so the SL interference must also be controlled. Virtual circuit routing

is considered from BS to UEs in the cell. The objective is to minimize the aggregate SL energies expended on all RBs while routing the transmissions from the BS to all UEs in the network. Downlink SNR and SL SINR thresholds must be satisfied for all RBs. The problem is an MINLP with a bilinear mixed integer objective function.

The RBMP is an LP in the power decision variables that grows by one additional constraint in each GBD iteration. Dual simplex and sensitivity analysis are used to exploit this by using the RBMP's solution from the previous iteration to reduce the computational effort resulting from its growing size. After necessary equivalent reformulations, the auxiliary sub-problem turned out to be an LP. When partial Lagrangian is applied to the sub-problem, it results in a linear minimum cost network flow program (MCNFP) structure and another trivial mixed binary linear problem with box constraints only. The network simplex is used to solve the resulting MCNFP. In case the primal sub-problem is infeasible, an l_1-norm feasibility sub-problem is solved.

B. GBD for Load Optimization in Cellular and Public WLAN Interworking Network

In [13], public WLAN is exploited to expand the capacity of existing cellular access networks. A main problem that is addressed is the underutlization of the public WLAN APs. For the amelioration of the traffic load, a centralized optimization model that jointly optimizes UE-access node (AN) association and resource allocation to maximize the overall throughput with some throughput fairness considerations, is proposed. Users can access only one AN simultaneously, either a cellular BS or a WLAN AP, depending on the strength of the received power. The association is represented by binary decision variables, while the resource share of the cellular network BS is modeled by a continuous variable. The objective function is mixed integer quadratic in the decision variables of the UE-AN association and the BS resource ratio share.

The problem is mixed integer quadratic program. The RBMP is a pure binary linear program, which is responsible for the UE-AN association. In each iteration, a tentative

TABLE III: Summary of design problems in which GBD is employed.

Network	Design problem	Objective	Decision variables	RMBP	Primal/ feasibility sub-problem	Computational performance
• Interworking public WLAN and cellular networks [15]	• UE-access node (AN) association and resource allocation for interworking networks' load balancing	• Maximize aggregate interworking network throughput	• Binary association decision variables • Continuous resource ratio decision variables	• Pure binary linear program, responsible for the UE-AN association	• Resource allocation quadratic program	• 107 iterations in average to convergence • Outperforms simple branch and bound, especially as the number of UEs increases
• D2D networks [13]	• Power control, interference management and routing	• Minimize the aggregate D2D relaying energy	• Binary selection variables indicating the links used for routing • Integer decision variables indicating the number of UEs whose traffic are routed on a given link • Continuous variables indicating the powers on D2D links and downlink	• Continuous LP that determines the powers in the network and manages the interference	• Linear minimum cost network flow problem that performs routing over multihop D2D relaying	• 15 iterations in average for convergence
• D2D networks [14]	• Power and channel allocation for D2D multicast groups	• Maximize the aggregate rate of all cellular users and D2D groups	• Binary variables for channel allocation • Continuous variables for power assignment	• Mixed binary linear program responsible for assignment of channels	• Convex non-linear program that performs power assignment and SINR satisfaction	• 5 iterations in average for convergence

association is tested for feasibility in a continuous non-linear quadratic sub-problem, which generates an optimality cut, if feasible. If this is not feasible, a cut of the type in (1) is used as a feasibility cut.

C. GBD for Joint Power Control and Channel Allocation in Multicast D2D Networks

Multicast D2D communications underlaying a cellular network is investigated in [14]. In this problem, each D2D group can reuse the uplink channels of multiple cellular users, and the channel of each cellular user can be reused by multiple D2D groups. A variant of GBD is used for solving an MINLP to find a joint power control and channel allocation solution that maximizes the aggregate rate of all cellular users and D2D groups. Minimum SINR requirements are imposed by constraints. The transmission powers are continuous variables and the channel allocations are modeled by binary variables. The MINLP is decomposed into a sub-problem and an RBMP: the former is a convex NLP that corresponds to the original problem with fixed binary variables, and the latter is a mixed binary LP derived through nonlinear duality theory using Lagrange multipliers obtained from the former. The auxiliary sub-problem is responsible for the power allocation and satisfaction of SINR requirements, while the RBMP is responsible for the subchannel assignments. In case the auxiliary sub-problem is infeasible, an l_1-norm feasibility sub-problem is solved.

VII. CHALLENGES AND FUTURE DIRECTIONS

The aforementioned variants of BD provide vast capabilities in developing a tractable solution methodology for different design problems in various communication networks. Challenges in BD include [1]: time-consuming iterations; poor feasibility and optimality cuts; low effectiveness of first few iterations; zigzagging primal solutions; slow convergence at a large number of iterations near the optimal solutions; upper bounds that do not improve for many iterations because multiple solutions with the same objective function value exist.

Improving the convergence can be obtained by reducing the number of iterations and the time required per iteration [1]. The former can be achieved by improving both the quality of the generated solutions and the cuts. The later can be obtained by exploiting structures in the MP and the auxiliary sub-problem. In what follows, we briefly discuss some of the possible enhancement strategies for BD.

A. Decomposition Strategy

This determines how the problem is projected to get the initial MP and the auxiliary sub-problem(s). All the linking constraints and non-complicating variables are projected out in the *conventional* decomposition. A discussion was given in [15] on how this causes the MP to lose all the information associated with non-complicating variables. As a result, there can be instability, a large number of iterations, and irregular progression of bounds. *Partial BD* strategies were proposed in [15] that add to the master explicit information from the sub-problems, either by retaining or creating scenarios, or even both. Significant improvements in terms of computational time and number of generated cuts were obtained. Researchers designing decision-making algorithms in communication networks can explore ways to exploit this, especially if the problem is a stochastic program.

B. Solution Procedure

This is specifically related to the algorithms used for the MP and the auxiliary sub-problem(s). The common standard techniques would be the simplex and branch-and-bound algorithms, which are usually treated as blackbox solvers. Advanced techniques make use of the structure of the MP and the auxiliary sub-problem or the search requirements of the BD algorithm. It was reported that 90% of the BD computational effort is needed for solving the MP.

Size management of the MP can be done by removing unnecessary cuts or avoiding the insertion of available cuts. Efficient solution methods for the MP include:

- Solving the MP to ϵ-optimality, especially at the early iterations of the BD.

- Solving the MP using meta-heuristics.
- Using *constraint programming* to handle special constraints more efficiently.
- Using *column generation* to handle certain structures effectively.

For the sub-problem, specialized algorithms can be used (e.g., transportation simplex, Hungarian algorithm, second-order cone programming) if it has a special structure.

C. Solution Generation

This has to do with the method for setting the trial values of the complicating variables. *Heuristics*, *alternative MP* or an *improved MP* can be exploited to generate solutions more quickly or for obtaining better solutions. *Hybrid* approaches can also be used. Even simple heuristics can generate high-quality initial solutions and cuts, repair infeasible solutions and reduce the computational cost of the MP and sub-problem. Finally, altering the MP, even temporarily, can mitigate slow generation of solutions (possibly with low qualities) and instability.

D. Cut Generation

This is mainly related to the technique used for the generation of optimal and feasibility cuts. The conventional approach does so by solving the auxiliary (or feasibility) sub-problems obtained from the decomposition. This can be inefficient for solving certain problems where the sub-problems are degenerate. Therefore, other possible strategies would invoke reformulations of the sub-problems.

VIII. Conclusion

This article reviewed the main variants of BD with existing applications to various communications networks. The BD basics were introduced and the impact of formulation on the computational performance was discussed, along with the ties of BD to other existing decomposition schemes. Each BD variant was presented along with with some of its existing applications including D2D networks, WSNs, WLANs, and infrastructure wired networks. Existing challenges in BD and ways in which these can be addressed were also discussed. In a nutshell, BD is a powerful technique that can find numerous applications to various network design problems.

IX. Acknowledgment

References

[1] R. Rahmaniani et al., "The Benders decomposition algorithm: A literature review," *European Journal of Operational Research*, vol. 259, no. 3, pp. 801-817, Jun. 2017.
[2] F. J. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische Mathematik*, pp. 238-252, Dec. 1962.
[3] G. Codato, and M. Fischetti. "Combinatorial Benders' cuts for mixed-integer linear programming," *Operations Research*, vol. 54, no. 4, pp. 756-766, Aug. 2006.
[4] B. Gendron et al., "Benders decomposition for a location-design problem in green wireless local area networks," *Electronic Notes in Discrete Mathematics*, vol. 41, pp. 367-374, Jun. 2013.
[5] C. Wolf and A. Koberstein, "Dynamic sequencing and cut consolidation for the parallel hybrid-cut nested L-shaped method,"*European Journal of Operational Research*, vol. 230, no. 1, pp. 143-156, Oct. 2013.
[6] A. M. Costa, "A survey on benders decomposition applied to fixed-charge network design problems," *Computers & Operations Research*, vol. 32, no. 6, pp. 1429-1450, Jun. 2005.
[7] L. P. Qian, Y. J. A. Zhang, Y. Wu, and J. Chen, "Joint base station association and power control via Benders' decomposition," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1651-1665, Apr. 2013.
[8] J. Zhang, E. Modiano, and D. Hay, "Enhancing network robustness via shielding," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2209-2222, Aug. 2017.
[9] H. Lin and H. Üster, "Exact and heuristic algorithms for data-gathering cluster-based wireless sensor network design problem," *IEEE/ACM Trans. Netw.*, vol. 22, no. 3, pp. 903-916, Jun. 2014.
[10] B. Behdani, J. C. Smith, and Y. Xia, "The lifetime maximization problem in wireless sensor networks with a mobile sink: mixed-integer programming formulations and algorithms," *IIE Transactions*, vol. 45, no. 10, 1094-1113, Oct. 2013.
[11] R. Kaewpuang, D. Niyato, and P. Wang, "Decomposition of stochastic power management for wireless base station in smart grid," *IEEE Commun. Lett.*, vol. 1, no. 2, pp. 97-100, Apr. 2012.
[12] A. Ibrahim, T. M. N. Ngatched, and O. A. Dobre, "Using Bender's decomposition for optimal power control and routing in multihop d2d cellular systems," accepted for publication in *IEEE Trans. Wireless Commun.*, Jul. 2019.
[13] F. Zhou, W. Li, L. Feng, P. Yu, and L. Meng, "Generalised Benders decomposition-based load optimisation in cellular and public WLAN interworking network," *IET Communications*, vol. 11, no. 12, pp. 1823-1830, Aug. 2017.
[14] H. Meshgi, D. Zhao, and R. Zheng, "Optimal resource allocation in multicast device-to-device communications underlaying LTE networks," *IEEE Trans. Veh. Technol*, vol. 66, no. 9, pp. 8357-8371, Sep. 2017.
[15] T. G. Crainic, M. Hewitt, and W. Rei "Partial decomposition strategies for two-stage stochastic integer programs," *Centre Interuniversitaire de Recherche sur les Réseaux d'Enterprise, la Logistique et le Transport (CIRRELT)*, Mar. 2014.

Ahmed Ibrahim (M'19) is currently a post-doctoral fellow and a per course instructor with the Memorial University of Newfoundland. He received his Ph.D. degree in electrical and computer engineering from the University of Manitoba (UoM) in October 2016. He coauthored a book entitled Optimization Methods for User Admissions and Radio Resource Allocation for Multicasting over High Altitude Platforms published by River Publishers in February 2019. He was a recipient of the Best Paper Award from the IEEE Wireless Communications and Networking Conference (WCNC) in 2019. He was awarded as an Exemplary Reviewer of the IEEE COMMUNICATIONS LETTERS journal in 2017.

Octavia A Dobre (M'05–SM'07) is a Professor and Research Chair at Memorial University, Canada. She was a Visiting Professor at Massachusetts Institute of Technology, as well as a Royal Society and a Fulbright Scholar. Her research interests include technologies for 5G and beyond, as well as optical and underwater communications. She published over 250 referred papers in these areas. Dr. Dobre serves as the Editor-in-Chief (EiC) of the IEEE Open Journal of the Communications Society. She was the EiC of the IEEE Communications Letters, a senior editor and an editor with prestigious journals, as well as General Chair and Technical Co-Chair of flagship conferences in her area of expertise. She is a Distinguished Lecturer of the IEEE Communications Society and a fellow of the Engineering Institute of Canada.

Telex M. N. Ngatched (M'05–SM'17) is an Associate Professor and Coordinator of the Engineering One program at the Grenfell Campus at Memorial University, Canada. His research interest include 5G enabling technologies, cognitive radio networks, visible light and power-line and communications, optical communications for OTN, and underwater communications. Dr. Ngatched is an Associate Editor of the IEEE Open Journal of the Communications Society and the Managing Editor of IEEE Communications Letters. He was the publication chair of IEEE CWIT 2015, an Associate Editor with the IEEE Communications Letters from 2015 to 2019, and Technical Program Committee (TPC) member and session chair for many prominent IEEE conferences in his area of expertise.

Ana Garcia Armada (S'96–A'98–M'00–SM'08) is a Professor at University Carlos III of Madrid, Spain. She has published around 150 referred papers and she holds four patents. She serves on the editorial board of IEEE Communications Letters, IEEE Trans. on Communications and IEEE Open Journal of the Communications Society. She has served on the TPC of more than 50 conferences and she has been part of many organizing committees. She has received several awards from University Carlos III of Madrid, including an excellent young researcher award and an award to best practices in teaching. She was awarded the third place Bell Labs Prize 2014 for shaping the future of information and communications technology. Her research mainly focuses on signal processing applied to wireless communications.
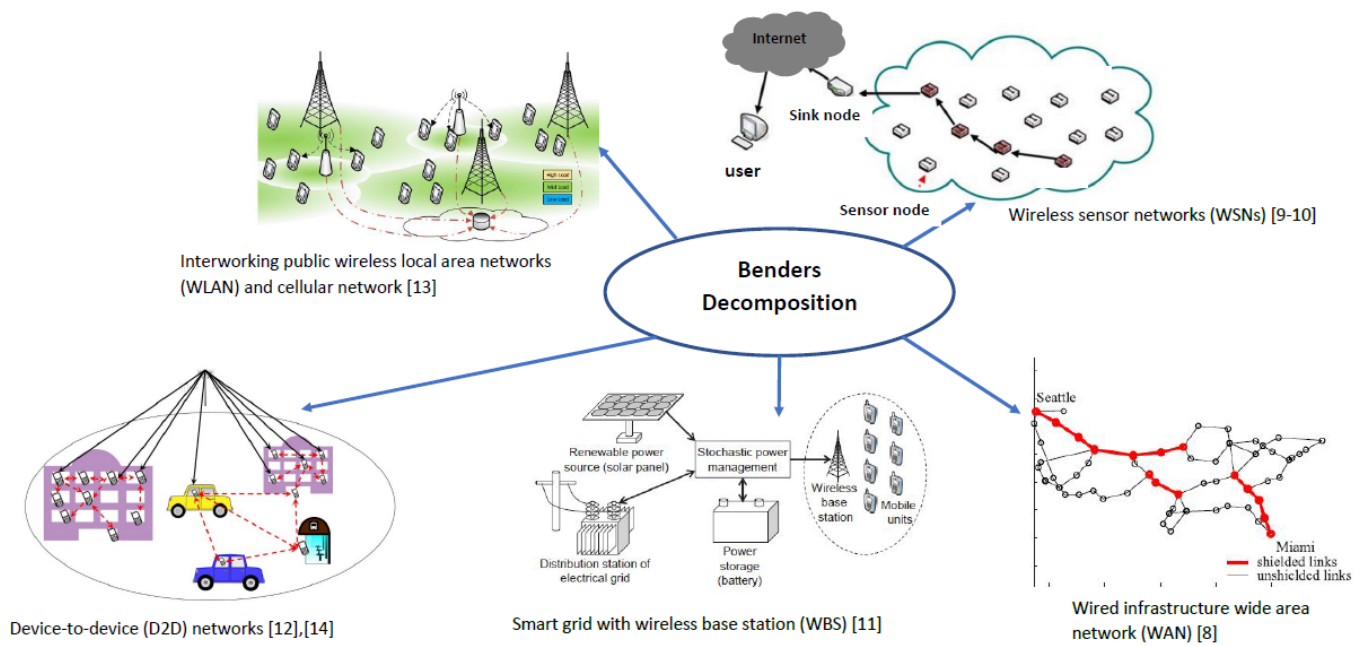
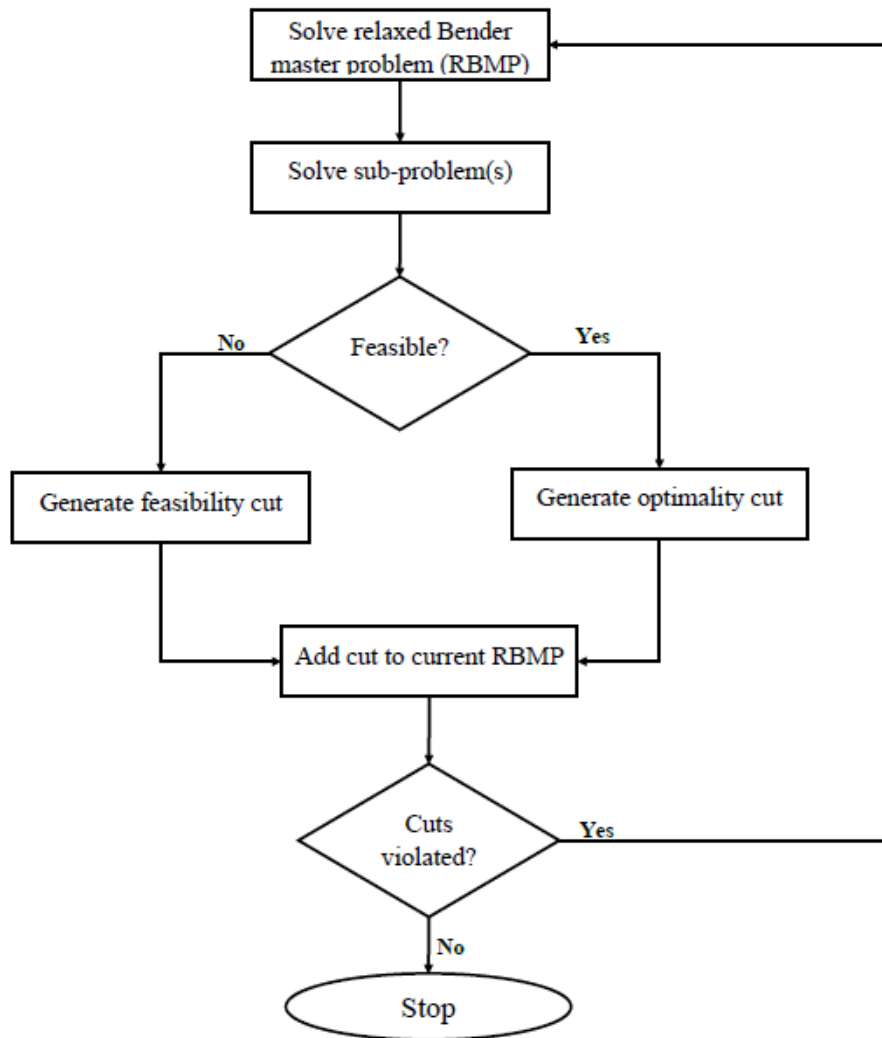Figure 1: Different network types in which BD applications exist.

Figure 2: Flowchart representation of BD.

Network	Design problem	Objective	Decision variables	Enhancements in CBD	RMBP	Sub-problem	Computational performance
• Heterogenous multicellular network [7]	• Joint user association and power control	• Throughput maximization and power consumption minimization in downlink	• Binary association decision variable • Continuous power decision variables	• Relaxation of integer requirements • Use of feasibility pump heuristic to restore feasibility	• User to BS association	• Power allocation LP	• **BS user association and power control (BARN):** ○ 51 CBD Iterations • **Accelerated-BARN:** ○ 16 CBD iterations ○ 30% lower CPU time than BARN
• Wired infrastructure WAN [8]	• Shielding links to ensure connectivity of all nodes in the occurrence of failures	• Minimize the costs needed for shielding	• Binary selection variables indicating whether a link is shielded • Binary variables indicating whether a link is destroyed for a particular failure event • Continuous variables indicating the flow on a given link	• Multiple failures are checked instead of one	• Determines which links to shield	• A zero objective optimization linear network flow problem obtains a feasible flow for a corresponding shielding solution	• Modified BD is faster than the classical BD by 50 %
• WSNs [9]	• Optimal cluster head, sink nodes selection and flow routing in multi-sink data-gathering WSNs	• Maximize the lifetime of the network	• Binary variables for selecting cluster head nodes and sink nodes • Continuous variables for assigning routing flows	• Upper bound heuristic • Strengthened cuts • ϵ optimal approach	• Gives a tentative selection of cluster head and sink locations	• Determines flow routing and sensor energies	• Enhanced CBD outperforms CPLEX branch and cut significantly for all instance sizes, with about 90% on average in the largest instance experimented, at an optimality gap of 2%
• WSNs [10]	• Lifetime maximization problem in WSNs with a mobile sink	• Maximize the lifetime of the network	• Locations the sink should visit (binary) • Order with which the sink should visit a location (binary) • Flow volume on each link (continuous) • Time the sink must spend at a given location (continuous)	• N/A	• Determines the sink locations to be visited and their order	• Time duration the mobile sink should stay at a location • Flow assignment between two sensors	• Significant average CPU time savings as compared to direct solve by CPLEX

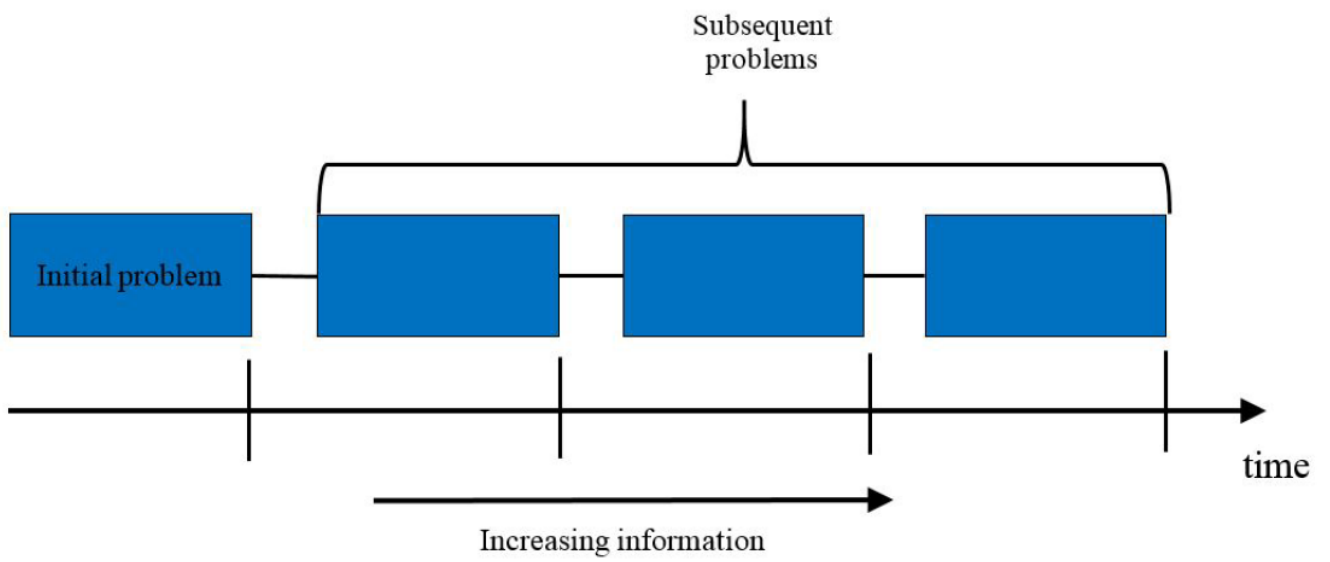TableII(a): Summary of design problems

Figure 3(a): Deterministic dynamic problem: There is no uncertainty about the evolution of information.
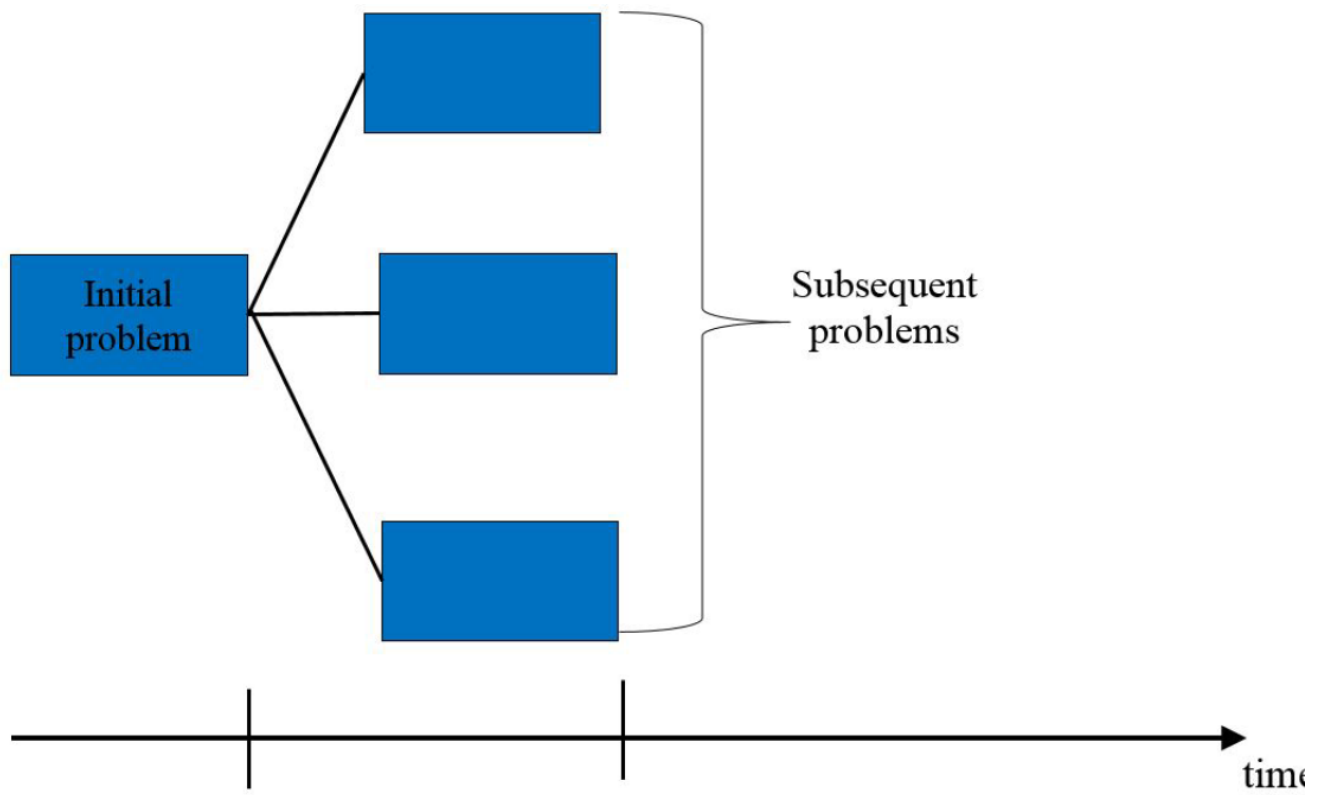
Figure 3 (b) Two-stage stochastic dynamic problem. Here there is uncertainty about which of the three possible successor states the problem moves into after the first decision, such that the decision must consider the uncertainty.

Network	Design problem	Objective	Decision variables	RMBP	Primal/ feasibility sub-problem	Computational performance
• Interworking public WLAN and cellular networks [15]	• UE-access node (AN) association and resource allocation for interworking networks' load balancing	• Maximize aggregate interworking network throughput	• Binary association decision variables • Continuous resource ratio decision variables	• Pure binary linear program, responsible for the UE-AN association	• Resource allocation quadratic program	• 107 iterations in average to convergence • Outperforms simple branch and bound, especially as the number of UEs increases
• D2D networks [13]	• Power control, interference management and routing	• Minimize the aggregate D2D relaying energy	• Binary selection variables indicating the links used for routing • Integer decision variables indicating the number of UEs whose traffic are routed on a given link • Continuous variables indicating the powers on D2D links and downlink	• Continuous LP that determines the powers in the network and manages the interference	• Linear minimum cost network flow problem that performs routing over multihop D2D relaying	• 15 iterations in average for convergence
• D2D networks [14]	• Power and channel allocation for D2D multicast groups	• Maximize the aggregate rate of all cellular users and D2D groups	• Binary variables for channel allocation • Continuous variables for power assignment	• Mixed binary linear program responsible for assignment of channels	• Convex non-linear program that performs power assignment and SINR satisfaction	• 5 iterations in average for convergence

TABLE III: Summary of design problems in which GBD is employed.