

Time-Dependence in Multi-Agent MDP Applied to Gate Assignment Problem

Oussama AOUN, Abdellatif EL AFIA
Operations Research and Logistics Team
ENSIAS Mohammed V University
Rabat, Morocco

Abstract—Many disturbances can impact gate assignments in daily operations of an airport. Gate Assignment Problem (GAP) is the main task of an airport to ensure smooth flight-to-Gate assignment managing all disturbances. Or, flights schedule often undergoes some unplanned disruptions, such as weather conditions, gate availability or simply a delay that usually arises. A good plan to GAP should manage as possible stochastic events and include all in the planning of assignment. To build a robust model taking in account eventual planning disorder, a dynamic stochastic vision based on Markov Decision Process theory is designed. In this approach, gates are perceived as collaborative agents seeking to accomplish a specific set of flights assignment tasks as provided by a centralized controller. Multi-agent reasoning is then coupled with time dependence aptitude with both time-dependent action durations and stochastic state transitions. This reflection will enable setting up a new model for the GAP powered by a Time-dependent Multi-Agent Markov Decision Processes (TMMDP). The use of this model can provide to controllers at the airport a robust prior solution in every time sequence rather than bringing a risk of online schedule adjustments to handle uncertainty. The solution of this model is a set of optimal decisions time valued to be made in each case of traffic disruption and at every moment.

Keywords—Time-dependent Multi-Agent Markov Decision Processes; stochastic programming; flight delays; Gate Assignment Problem

I. INTRODUCTION

More interest in recent years is allowed to providing advanced techniques in the air traffic framework. This is resulting from the increase of air transport traffic [1]. The main objectives are best allocation and management of airport and airline resources in the best way effectively and efficiently. Caused by the dynamic stochastic operational environment conditions of air transport, the scheduling problems currently confronted by the airport and airline managers are leading to challenging and complex planning problems that involve innovative models and solutions. This is triggered by the significant diversity of resource segments that have to be regarded including terminals, flights, crews, baggage ..., and most are interdependent. In fact, stochastic disruptions in air traffic transport raised the complexity of the resolution models. This is progressively more taken under consideration in most recent studies.

The main target of an airport is to guarantee a fluent flights traffic. Optimal assignment of aircraft guaranteed to make available over time the proper gates. If an aircraft is not

assigned, it will be forced to wait on the ramp very well as in the air; This type of scenarios are quite undesirable on account of time wasting and let to flight delays. Also, ramps and airspace are as well resources with a limited capacity.

Gate flight assignment is an essential task of an airport; it is the primary activity in airline traffic transport management [2]. Moreover, several airports today have severe capacity constraints resulting from the increase in air traffic volume. The GAP can be regarded as such a problem of constraint resource assignment, in which gates represent resources and aircraft considered as resource consumers.

Furthermore, GAP is thought to be a challenging problem [3] since it includes very inter-dependent resources integrating aircraft, crews, and gates. Therefore, severe disruptions in the airport manifested as flight delays are caused by inadequate assignment, which reduces the customer services and produces inefficient use of gate services and conflicting flights.

Various circumstances could potentially cause stochastic disruptions in gate assignment; it can possibly be interrelated to possible gate dysfunction, a flight delay or earliness, extreme weather conditions, or for any more causes. This type of daily disturbances might reduce the overall performance of the currently assigned gates once associated with actual operations. Therefore, even a unique variation in a single flight plan could engender a series of disturbances for additional aircraft, which have been designated to the same gate. This sort of phenomenon is very unwanted in airline operations due to its noticeable costly impact.

Various GAP models and techniques are identified from the literature. Static as well as stochastic models are developed. Working with methods with an exact solution can be obviously more suitable. However, [4] states that these kinds of exact methods are actually ineffective to resolve real problems. This is because flights in static models are allocated to gates depending on the expected flight schedule using fixed parameters. Nonetheless, in real operations, stochastic disruptions occur frequently, leading to real-time adjustments of gate assignments and flight delays. Consequently, stochastic methods have been widely motivated in recent researches.

Consequently, to build a significantly better gate flight assignment approach, it has to include in the model the possibilities of stochastic flight delays that may arise in real operations.

When it comes to stochastic environments, Markov Decision Processes (MDPs) [5] have confirmed to be effective in optimal decision making. A derived version of MDPs called multi-agent Markov decision process [6] was developed to manage some challenges in the standard GAP based MDP firstly introduced. In this work, collaborative multi-agent based MDPs is built, which are composed of multiple agents attempting to produce the best allocation of aircraft to gates. A new methodology for GAP is provided and regarded as a multi-agent problem that includes robustness for a stochastic disturbance. The GAP is therefore designed as a Multi-agent MDP that is intended to resolve within the assumption of environment uncertainty the GAP. Then, incorporating Time dependence to the developed Multi-Agent model enables further stochastic planning ability. In this method, the stochastic feature is considered caused by flight delays with the flexibility to consider additional constraints for the constructed model. Gates are designed as agents having a centralized controller. Consequently, individual agent or gate possesses full visibility of airport operations and so can be aware of flights allocated to every gate at the time of planning horizon. Built policies take into consideration the time dimension. Time-dependent Multi-agent Markov Decision Processes allows more real illustration of the Gate problem with rewards and transitions varying with time. So TMMDP includes Multi agent aspect coupled with a real-valued time component.

This paper is structured as follows: the next section provides a literature review. Section 3 shows the MMDP model for the GAP. Section 4 presents the model of GAP in stochastic circumstances with time dependence. Furthermore, Section 5 will provide experimentations, and at last, conclusion and perspective are dressed.

Airport Gate Assignment Problem is referred to as setting an appropriate gate for each arriving aircraft to the airport until the time of its departure. It is one among the primary components in what concerns the management of airport resources. Gates, being a resource, is subject to the next two groups of constraints as categorized in literature: strict and soft constraints (see [3]).

The first category is obligatory to represent the problem of gate assignment. It comprises these constraints:

- Single: Each aircraft have to be assigned just to one gate.
- Feasible: A single gate could be assigned to one single aircraft simultaneously.

Soft constraints are various and can possibly be related to either airlines or airports. Mainly common among constraints in the literature is about to minimize the total walking distances within passenger transfer. (e.g. [7]), or just like assigning aircraft to some specified gates, also taking into consideration the size aircraft for allocating the gate [8]. It can also be minimizing the number of aircraft obligated to wait for a gate.

There are several objective functions in GAP models. Some notable ones of literature are cited. These functions are like minimizing the total walking distance [7] or the total

waiting time for passengers such as in [9] or also minimizing the number of un-gated aircraft in [2]. Others like minimizing the current schedule modification from an initial schedule, or also maximizing the preferences of assigning particular aircraft to individual gates (e.g. [10]) and minimizing gate conflict in [2]. In this paper, the stochastic model will implement particularly the last one to minimize conflicting assignment due to flights disruptions.

The GAP formulation is classified into two main types: deterministic and stochastic models. In the first kind, just static parameters are regarded (including passengers, gates, number of flights...); due to stochastic perturbations in real-world operations deterministic models becomes infeasible. Stochastic GAP models have been investigated to consider those disruptions in air traffic into concern such as flight delays or some severe weather conditions.

Deterministic models are more a lot discussed in the literature, such as [7]. Most have as an objective the minimization of the total passenger-walking distance. Lately, stochastic and robust models are more reviewed assisting operators to act in response to possible uncertain events.

To illustrate stochastic and robust GAP resolutions in literature, [11] displays that having a planned buffer time into the flight schedule can increase schedule punctuality. In [12] and [9], they use in their GAP a fixed buffer time among two consecutive flights assigned to the same gate in order to absorb the possible stochastic flight delays. In [12] author produces a multi-commodity network flow approach as well as in [13]. In [14], author builds up a heuristic approach sensitive to stochastic flight delays in a framework that consists of three components, a stochastic gate assignment model, then a real-time assignment rule, in addition to two penalty correction methods.

In [2], GAP is modelled as a stochastic programming model and altered it into a binary programming model; the resolution contains hybrid meta-heuristic, a tabu search, and a local search. Also, an ant system combined with a local search in [15] has been used to an over-constrained airport Gate Assignment Problem with the interest of choosing and allocating aircraft to the gates minimizing the total passenger interconnection.

Recently, a model based heuristics of Mixed Integer Programming in [16] has been presented, it has been confirmed to be more efficient when compared to the linearized models, and more robust. Likewise, a multi-objective optimization model of GAP has been offered in [17], a particle swarm algorithm for resolution is used for resolution, which gives an improved comprehensive service of gate assignment regarding robustness. Applying also a metaheuristic for resolution, authors in [18] designed a three-objective problem to the GAP and using a non-dominated sorting genetic algorithm for resolution.

Markov process theory, in general, has been proven for application in airline transport like in [19]. Notably, the use of MDP model for GAP has been applied in [20] to deal with gate disturbances with consideration of aircraft size in the assignment, where neighboring gates can just only accept

aircraft of a specified size or are possibly blocked. A most recent robust GAP with multi-agent MDP model has been provided in [21].

In a similar idea of incorporating stochastic disturbance for establishing gate assignment, a multi-agent system with time dependence for modeling with time dependence is used in this paper. Multi-agent systems (SMA) are a part of Distributed Artificial Intelligence. Their applications are large: game theory, humanities, economics, and other real-world applications including air traffic control, robotics, and networking. SMA methods are interested in connections between independent entities. This circumstance is mainly examined in SMA as the cooperation that requires complex components.

In planning with multi-agent systems, it is commonly supposed to possess some number of agents, each one with their individual group of actions, and a provided tasks to be solved recognizing that interaction with the other agents is essential. Reinforcement learning has been a practical methodology to construct coexisting agents (e.g. [22]) as well as Markov games (see, e.g. [23]). In general, each agent may possess its personal goals. In this paper, the concern is given to the case of fully cooperative agents; where all of the agents have a shared similar goal to maximize the total expected reward. In particular, where agents are autonomous and distributed, a local Markov Decision Process (MDP's [5]) is used to express every single agent's state and actions space. Therefore, the utility of any given system state is similar for all agents, and with models of uncertainty and general utility, Multi-agent Markov decision process (MMDP) is developed by [24] to incorporate such numerous adaptive agents that interact to compute some given goals. MMDP has been applied in various domains as well as in the air transportation (see [25]).

MMDP is the basis of full observability of the global state by every single agent; it is designed as a set of interacting learners agents, which are autonomous. These agents have to learn in order to cooperate and obtain their assigned goal. It can also either centralized or decentralized in term of decision-making main feature [26]. Hence, this paper incorporates Markov decision processes as a formalism in the multi-agent structure (e.g. [24]). It supposes having a centralized controller knowing all information regarding the system (Fig. 1), including actions, the global state of the system, and rewards; thus the controller possesses the decision authority and keeps information distributed among agents.

Multi-Agent notion can as well be combined with real-time valued to include time evolution into the multi-agent system dynamics. A Time-dependent Markov Decision Process (TMDP) is provided by [27] to give this extension. This model is composed of stochastic state transitions and as well as stochastic time-dependent action durations. The actions in TMDP model are stochastic and time-varying:

$$a(t) \sim \text{policy}(s, a(t)) \quad (1)$$

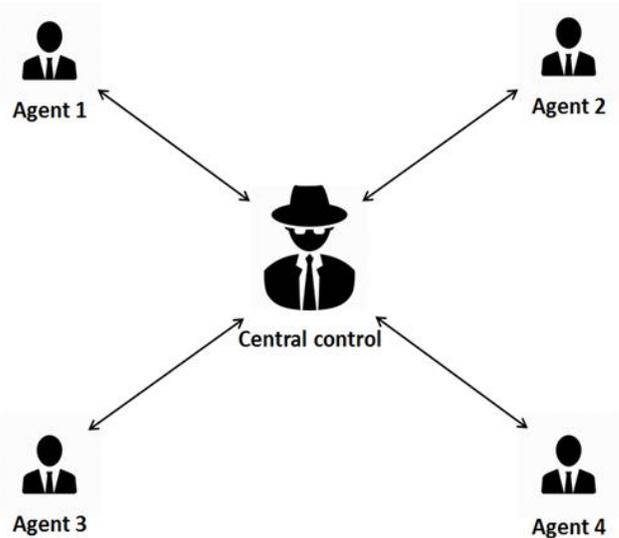


Fig. 1. Centralized control in MMDP.

Resulting policies are actions to be performed by agents in every single time sequence. Then, the real planning window can be widespread to problems under uncertainty changing with time.

So, in this formulation as in [28], first, MMDPs consider an assignment centered decomposition approach, which is intermediate between the join MDP method and the method of independent agents. The centralized controller is adopted having the complete relevant information regarding the states of all agents to allocate jobs and assign jobs and resources to agents determined by a task level value functions associated with agents. After the jobs are allocated to agents, the particular lower level actions of agents are driven by the task level value functions till the primary controller reassigns jobs. Then, adding time dependence behavior will give a more realistic representation of the Gate Assignment Problem, inspired by TMDP and coupled with the MMDP approach providing a new formalism of time-dependent Multi agent MDP. This method will help us to have real-time policies to apply in every case of disturbance for the GAP problem.

II. THEORETICAL BACKGROUND

Giving the theoretical knowledge, Markov Decision Processes (MDPs) are defined (see [5]), and then generalized to multi-agent settings. Then, the basic model of Time-Dependent Markov Decision Process (TMDP) (given by [27]) is provided to finally conclude a new extension of MMDP depending on time and formalize the Time-Dependent Multi-Agent Markov Decision Process (TMMDP).

A. Standard Markov Decision Process

Considerably, more research interested in problems having uncertainty in the planning with possibly conflicting objectives. As a tool of artificial intelligence (AI) planning, decision-theoretic dress those challenges, especially, Markov Decision Processes Theory (MDPs). It finds a significant attractiveness in recent researches equally as a computational and conceptual model. MDP is defined by a tuple $\langle S, A, P, R \rangle$ where s is a finite set of states S describing

systematic interests, a finite set of actions A featured to the agent, and then a reward function R . When an action can take an agent from one state to a second one, the results of actions is uncertainty described by the probability P considered as transition model. A mapping $\pi : S \rightarrow A$ defines a policy. The objective is to identify the optimal policy π^* maximizing per each state the expected discounted future reward. MDP is considered in this paper to possess an infinite horizon with exponentially discounted future rewards by a discount factor $\gamma \in [0, 1)$.

B. Multi-Agent Markov Decision Processes

The MDP model can be extended to multi-agent systems to define The Markov Decision Processes Multi-agent or MMDP as in [6]. In this formalism, the same goal of maximizing the total expected reward is shared for all agents having the same joint utility function. MMDP can be viewed as a generalization of MDP with a single agent; Or, but also a special case of Markov games [29] where the payoff function is identical for all agents. Let define first the MMDP formalism before offering it as a useful framework to constitute a new GAP model.

A MMDP is identified via a tuple $\langle n, S, A, P, R \rangle$. Where each one action is identified by the set of actions of all single agents, it constitutes a joint action. Each element is defined as :

- n : the total number of agents in the system.
- S : refers to the set of states S .
- $A = A_1 \times \dots \times A_n$: identifies the set of joint actions of all agents, A_{gi} defines the set of local actions designed for the agent i .
- P defines the transition function; it provides the probability $P(s, a, s')$ of the system moves from a state s into a state s' once agents run the joint action $a \in A$.
- R identifies the reward function. $R(s, a, s')$ is the reward received after moving from a state s to a state s' performing an action a .

Solving a MMDP is about determining a joint policy $\pi = \langle \pi_1, \dots, \pi_n \rangle$. Where π_i corresponds to the policy of a local agent. It identifies a function $\pi_i : S \rightarrow A_i$ that gives a mapping to any system state to the action of agent i . The joint policy will be computed applying the standard algorithm the Value Iteration (continue to operating in the general situation of decentralized agents, see [18]).

C. Time-Dependent Markov Decision Processes

In standard previously defined MDPs, transitions and rewards are thought to be stationary functions; they do not undergo any change during decision epochs. In literature, some approaches like [30] define Stochastic Time-Dependent Network where stochastic transition durations are included, but transition outcomes are deterministic. A model given by [27] is one of the first models to focus on time as an independent observable state variable; it is named as Time-dependent Markov Decision Process.

Time-dependent Markov Decision Process extends the Markov decision process model where a continuous

observable time dimension is contained in the state space. The added time variable allows more real representation of large problems with transitions or rewards time-varying. So TMDP includes problems with following properties:

- State transitions are stochastic;
- Time-dependent action durations are stochastic.
- Rewards are Time-dependent.

In the TMDP model, each transition, which arises from making an action, is decomposed into a set of possible outcomes $\{\mu\}$. Every single outcome identifies both a transition duration and a resulting state.

The TMDP model decomposes each transition resulting from the application of action into a set of possible outcomes $\{\mu\}$. Each outcome describes a resulting state and transition duration.

Formally, the TMDP is defined as in [27] by:

- S : Discrete space state.
- A : Discrete action space.
- M : Discrete set of outcomes, of the form $\mu = (s'_\mu, T_\mu, P_\mu)$:
 - $s'_\mu \in S$: is the resulting space
 - $T_\mu \in \{ABS, REL\}$: identifies the type of the resulting time distribution (if it is absolute or relative)
 - $P_\mu(t)$ (If $T_\mu = ABS$): probability density function (pdf) over absolute arrival times of μ
 - $P_\mu(\delta)$ (If $T_\mu = REL$): probability density function over durations of μ
- L : $L(\mu|s, t, a)$ is the likelihood of outcome μ given action a , state s , and time t
- R : $R(\mu, t, \delta)$ is the reward associated to outcome μ at time t with a duration δ

In the figure below (Fig. 2), it shows a simple graphic representation of TMDP evolution.

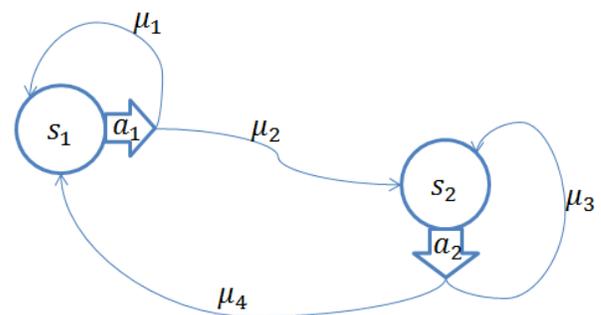


Fig. 2. Elementary example of TMDP.

In TDMDP and at time t , if in a state s_1 agent executes an action a_1 , it will be generated outcome μ_1 by certain probability $L(\mu_1|s_1, t, a_1)$ and an another outcome μ_2 by a probability $L(\mu_2|s_2, t, a_2)$. μ_2 represents the transition to s_2 and P_{μ_2} gives the transition absolute arrival time, while μ_1

represents the return to s_1 (failure to leave s_1) with a duration P_{μ_1} . Implicitly, a waiting time is inserted before each action in the model.

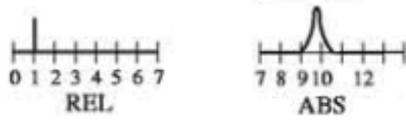


Fig. 3. Representation of pdf types.

The likelihood functions L governs possible outcomes in the model. Time distributions in a TMDP could be either “relative” (REL) or “absolute” (ABS) as shown as an example in Fig. 3.

The TMDP model can be represented by the Bellman equations below:

$$V(s, t) = \max_{a \in A} Q(s, t, a)$$

$$Q(s, t, a) = \sum_{\mu \in M} L(\mu|s, a, t) \cdot U(\mu, t)$$

$$U(\mu, t) = \int_{-\infty}^{\infty} P_{\mu}(t') [R(\mu, t, t' - t) + V(s'_{\mu}, t')] dt' \quad (2)$$

(if $T_{\mu} = ABS$)

$$U(\mu, t) = \int_{-\infty}^{\infty} P_{\mu}(t') [R(\mu, t, t' - t) + V(s'_{\mu}, t')] dt'$$

(if $T_{\mu} = REL$)

Where

$U(\mu, t)$: Utility associated to the outcome μ in time t

$V(s, t)$: Time-value function of the immediate action

$Q(s, t, a)$: Expected Q time-value through outcomes.

The resolution of this model is performed using Bellman equations, (2) representing an undiscounted continuous-time MDP. At each state, the optimal time-value function is a piecewise linear function of time, which could be precisely calculated by value iteration [27]. The TMDP model is more general than semi-Markov decision processes [31] that have no notion of absolute time. With absolute time included in the state space, comprehensive set of domain objectives can be modeled beyond the objective to minimize expected time, like for example the probability of designing a deadline. Actually, the variable time dimension may represent further quantities; it can consider planning with the non-linear utilities, or also with continuous resources.

D. Time-Dependent Multi-Agent Markov Decision Processes

Based on the two previous definitions of MMDP and TMDP, a new formalism is defined combining between those approaches. So, it is called Time-Dependent Multi-Agent Markov decision process TMMDP. This is a MMDP seen as cooperative multi-agent systems as in [6] or associated with a time dependence capabilities as defined by [27]. MMDP is then extended to take a continuous observable time dimension contained in the state space. Supposing time variable is

common between agents, a global time is associated to all agents.

A TMMDP is defined by:

- n : Number of agents.
- S : refers to the set of states s
- $A = A^1 \times \dots \times A^n$: The set of joint actions for the agents i is the set of local actions of the agent A_{g_i} .
- M : Discrete set of outcomes, of the form $\mu = (s'_{\mu}, T_{\mu}, P_{\mu})$:
 - $s'_{\mu} \in S$: the resulting space
 - $T_{\mu} \in \{ABS, REL\}$: identifies the type of the resulting time distribution (absolute or relative)
 - $P_{\mu}(t')$ (If $T_{\mu} = ABS$): pdf (probability density function) over absolute arrival times of μ
 - $P_{\mu}(\delta)$ (If $T_{\mu} = REL$): pdf over durations of μ
- L : $L(\mu|s, t, a)$ is the likelihood of outcome μ given join state s , time t and join action $a = (a_1, \dots, a_n)$.
- R : $R(\mu, t, \delta)$ Reward attached to outcome μ at time t for all agents with duration δ .

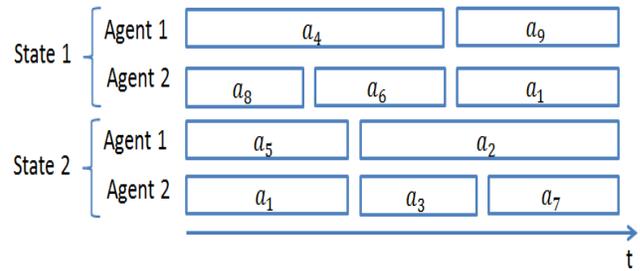


Fig. 4. TMMDP policy representation

The aim of defining TMMDP formalism is to model and solve large real problems of planning under uncertainty taking into account either cooperative agent property and time evolution. Resulting policies are actions to be performed by agents in every time sequence (see Fig. 4).

III. THE PROPOSED APPROACH

A. Multi-Agent Reasoning

Various efforts made in the literature to manage uncertainty (see Section 2). With the Objectif to build a robust, a multi-Agent based method is selected to develop a solution that can resist the most to flights disturbances. The choice for this specialized background to model the problem. MAS methods are getting large approval being an effective instrument to solve more complex problems and then designs a promising alternative. As well, many advantages related to multi-agent reasoning such as the distribution of processing, which made some type of problems more simple in conception. Additionally, it provides an intelligent alternative to complex problems and logical approach of decomposing into individual agents that cooperate.

In this paper, MAS is considered to be managed by a centralized controller, and the solution is composed of all possible decisions that could be taken within the planning horizon of gates to flights assignment. Therefore, This approach supposes there is no need to take real-time optimization since it is predetermined the solutions for all likely case of disturbances. Hence, for any provided gate assignment combination, the solution offers the best decision of gate allocation to make.

B. Time-Dependence Behavior

The real interest is given to sequential decision problems. Theoretical aspect based on MDPs gives a best well-known tool to model and solve them, giving optimal results. However, real-world problems have additional and specific behavior, which is time dependence. MDP reflects only fixed time steps between decision epochs, which can be easily modeled as iteration steps. This property does not reflect the real evolution of problems like the subject of gate assignment. To bypass this limitation, Time-dependent MDP (TMDP) has been proposed in those models (see the previous section), the transition between states is not instantaneous but proceeds in specific time t. Also in TMDP, the time is always observable, optimal policies give to the agent the best moment to make a decision or execute an action due to the state of the system.

Inspired by other occurrences like the truck dispatching system where decisions about truck assignments and destinations are made in real-time [32], choosing to benefit from temporal aspect and to project it to Gate Assignment Problem. Therefore, the rewards associated with action outcomes in the time-dependent frameworks will be represented as time-dependent functions including more real evolution information of the problem.

C. Multi-Agent Model for the GAP:

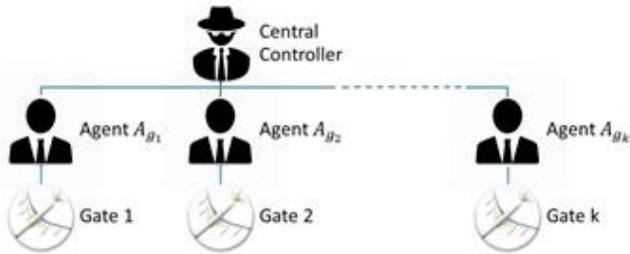


Fig. 5. Agents representation

Before extending the model of GAP to be time-dependent, an earlier formulation like in [21] of the Gate Assignment Problem with MMDP is presented. The model is given by is a tuple $\langle K, S, A, P, R \rangle$ as a follow (see Fig. 5):

The State $S = S_1 \times \dots \times S_K$ is a vector giving the diverse feasible combinations of flights indexed by its assignment position $S_i = (s_1, \dots, s_k)$, where k is number of gates and $s_i \in V$. V represents the set of flights to be allocated to gates during the planning horizon (one day in general).

The set of actions $A = A_1 \times \dots \times A_K$ describes the set of joint actions for the agents, A_i gives the set of local actions of

the agent i. For each single agent, performing a $a \in A_i$, will match an action of allocation a flight $a \in V$ to the gate i.

Therefore, each agent is in charge of handling a particular gate, and a $a \in A_i$ for agent i considers that there is a set of feasible flights to be affected to the gate i. $A_i \subset V$ that are appropriated to be allocated to gate i. This supposition regarded as a feasibility constraint that describes the possible assignment.

Defining: $A_{i,t}$ set of feasible flights for the gate i at a discrete time t. Then:

$$A_i = \sum_t A_{i,t} \quad , \quad (i \leq k) \quad (3)$$

P (s, s', a) gives the probability of transition as :

$$P: S \times S \times A \rightarrow [0, 1]$$

It represents the probability of the going from state s into another state s' when agents perform a joint action $\in A$. This probability is views as the possibility of modifying assignment combination from s to s' resulting from executing a re-assignment action.

The probability P is integrating the complete stochastic information about assignment of gates including stochastic delays as well as additional disturbances that impact gate assignment and computed as a probability of occurrence. This probability utilizes other estimation techniques to build the probabilistic model of GAP under possible disruptions.

The way how transition probabilities are defined is essential for building the robustness of the GAP based MMDP model. The state transition stochastic matrix P defines all likely possible state transition probabilities (p_{ij}):

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{bmatrix} \quad (4)$$

Where:

$$\sum_{j=1}^n p_{ij} = 1 \quad (i = 1, 2, \dots, n), p_{ij} \geq 0 \quad (i, j = 1, 2, \dots, n)$$

Various statistical estimating methods could be applied to calculate state transition probabilities described above. The method as in [33] is applied using statistical data of state transition. Actions corresponding to flights combination are identified, and the arising states are collected from data. The collected values from observed data, $k_1(a)$ corresponds to the case without disruption on state s_1 performing action a, and $k_{12}(a)$ is the case of disruption observed between state s_1 and state s_2 performing action a. therefore the transition probability between s_1 and s_2 performing an action a is estimated from observed data as :

$$p(s_1, s_2, a) = \frac{k_{12}(a)}{k_1(a)} \quad (5)$$

$R(s, a, s')$ corresponds to the reward acquired once transiting from a state s to a state s' performing an action a . This involves costs as negative reward or positive reward as benefits of each reassignment.

R is thought as : $R: S \times S \times A \rightarrow R$

Where its function is defined as:

$$R(s, a, s') = -\lambda \delta_{ss'} p(s, a, s') + \gamma(1 - \delta_{ss'}) p(s, a, s') \quad (6)$$

having:

- $\delta_{ss'} = 1$ if $s = s'$ and 0 otherwise
- λ Penalty unit
- γ Recompense unit

The main task of a decision maker is to compute a policy as:

$$\pi: S \rightarrow A$$

A state-action sequence of decisions that maximize the expected total reward is denoted as π^* , and corresponds to the policy optimal. $V^*(s)$ gives the maximum cumulative reward attained by the optimal policy beginning with states. Therefore, the optimal decision in a state s is to choose an action a maximizing the sum of the immediate reward $R(s, a, s')$ and the value V^* of the immediate successor state, discounted by γ ($0 \leq \gamma < 1$):

$$\pi^*(s) = \arg \max_a [R(s, a) + \gamma V^*(p(s, a))] \quad (7)$$

The solution concerns obtaining an optimal stationary policy π^* that maximize for each state s and for all agents the expected discounted future reward. π^* contains the optimal decisions to make in every gate considering the assignment state.

MMDP model representing the GAP problem is solved using the value iteration function determined by Howard algorithm (see [5]) and provided as follows:

Policy Iteration Algorithm [34]:
<ul style="list-style-type: none"> • π any policy • While $\pi \neq \pi'$ <ol style="list-style-type: none"> (1) $\pi := \pi'$ (2) For all $s \in S$ Compute $V_\pi(s)$ by solving the system of S unknowns given by Eq (1) (3) For all $s \in S$ If there exists an action $a \in A$ such that: $R(s, a) + \gamma \sum_{s' \in S} P(s, s', a) V_\pi(s') > V_\pi(s)$ Then $\pi'(s) := a$ Else $\pi'(s) := \pi(s)$ • Return π

This algorithm is assured to converge (as in [5]).

D. Multi-Agent with Time Dependence Model:

Based on the same approach as the previous model, this paper presents another model with Multi-Agent reasoning but

including the time evolution aspect of the Gate Assignment Problem. The considered Time-dependent Multi-agent Markov Decision Problem illustrated in Fig. 6.

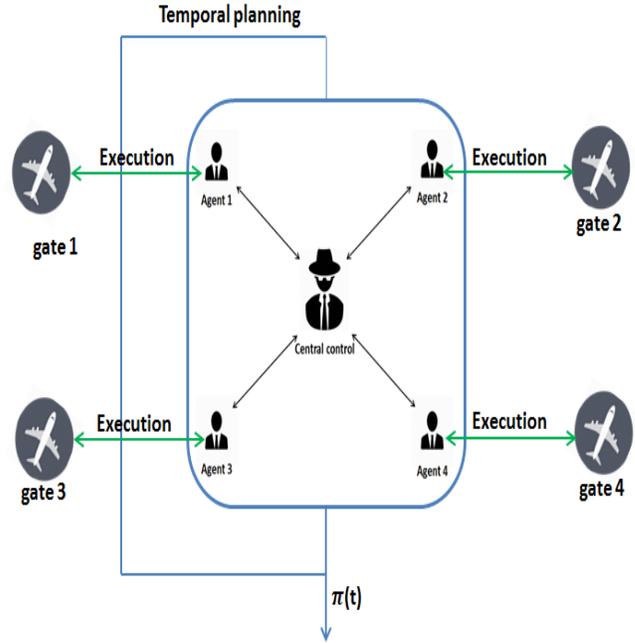


Fig. 6. Agents distribution and temporal planning

Let K be the Number of agents; also correspond to the number of gates. Taking the same actions definition from the previous model, the set of actions $A = A_1 \times \dots \times A_K$ defines the set of joint actions of agents being also for every agent i assigning a flight $a \in V$ to a gate I , V the set of flights.

Additional temporal information will be included first in the Discrete set M set of outcomes, of the form $\mu = (s'_\mu, T_\mu, P_\mu)$:

- $s'_\mu \in S$: the resulting state space
- $S = S_1 \times \dots \times S_K$ gives different possible combinations of flights $a \in V$.
- $T_\mu \in \{ABS, REL\}$: Type of the time distribution (absolute or relative).
 - If $T_\mu = ABS, P_\mu(t')$ will be a pdf over absolute arrival times of μ and corresponds to distribution time associated to some gates assignment configuration action.
 - If $T_\mu = REL, P_\mu(\delta)$: pdf will be over durations of μ that corresponds to the duration needed to establish the assignment configuration action.
- L : $L(\mu|s, t, a)$ is the likelihood of outcome μ given state of gate assignment s , time t and action of next assignment to execute $a = (a_1, \dots, a_n), a_i \in V$.
- R : $R(\mu, t, \delta)$ is the reward for the outcome μ at time t with duration δ , corresponding to reward of spending δ duration at time t with airport assignment action μ . The

reward includes as the previous model two components :

- A benefit from the gate assignment outcome μ .
- A penalty to assignment outcomes μ that causing a possible disturbance at time t and with duration δ .

IV. EXPERIMENT

A. Multi-Agent Model experiment:

Computational analysis is done to test the efficiency of the used Multi-Agent MDP approach, and utilizing a simple data example to conduct experimentations.

For simplification, data includes two gates and three aircraft to allocate in a discrete window of time between T_0 and T_3 .

$V_i \in V$, is set of flights and for $i = 0$ it match a vacant assignment gate.

As a sample, in this experimental instance exist three possible states:

$s_1 = (V_1, V_2)$, $s_2 = (V_3, V_0)$, $s_3 = (V_0, V_3)$. Two agents are affiliated to the two gates, therefore actions are: $a_1 = (V_1, V_2)$, $a_2 = (V_3, V_0)$, $a_3 = (V_0, V_3)$.

As an initial policy, the solution provided first by a deterministic approach to the problem from literature is used. Simple values are used as input parameters only for simulation. The preliminary policy is as follows: $\pi_0 = (a_2, a_1, a_1)$:

TABLE I. INITIAL POLICY WITHOUT DISRUPTION

		T_0	T_1	T_2
Agent 1	Gate 1	V_1		V_3
Agent 2	Gate 2	V_2		

It is designed regarding observations, transition probabilities and rewards are shown in Fig. 7.

>> P				>> R			
$P(:, :, 1) =$				$R(:, :, 1) =$			
1	0	0		-100	0	0	
1	0	0		100	0	0	
1	0	0		100	0	0	
$P(:, :, 2) =$				$R(:, :, 2) =$			
0.7500	0.2500	0		-75	25	0	
0	1.0000	0		0	-100	0	
0	0	1.0000		0	0	-100	
$P(:, :, 3) =$				$R(:, :, 3) =$			
0.0500	0	0.9500		-5	0	95	
0	1.0000	0		0	-100	0	
0	0	1.0000		0	0	-100	

Fig. 7. Transitions and rewards matrixes

With $\lambda = \gamma = 1$.

Like in Table I, $p(s_1, s_1, a_2) = 75\%$ expresses a probability of disruption performing action a_2 on s_1 , which

corresponds to the situation in Table II (V_1 is delayed and still allocated to gate 1 that V_3 cannot be re-assigned, which results in conflict):

TABLE II. CONFLICTING ASSIGNMENT IN INITIAL POLICY DUE TO DELAY

		T_0	T_1	T_2
Agent 1	Gate 1	V_1	Conflict	V_3
Agent 2	Gate 2	V_2		

A simple experimentation is done to demonstrate the feasibility of the suggested resolution method.

The initial policy is not possible as a result of delay of the flight V_1 (Table II), which causes a conflict in gate allocating. Therefore this solution is used as initial policy in the policy iteration algorithm then the algorithm is performed.

After execution of value iteration algorithm in MatLab, the provided solution offers another order in the gate assignment; optimal policy is $\pi^* = (a_3, a_1, a_1)$ identified as in Table III:

TABLE III. OPTIMAL POLICY

		T_0	T_1	T_2
Agent 1	Gate 1	V_1		
Agent 2	Gate 2	V_2		V_3

Table III shows that the proposed approach can give a solution that is more robust to delays. Compared with the sample agent MDP in [21], this approach is more representative of the problem structure because of the Multi-agent distribution of processing, that simplify its conception. Also, MMDP gives gate assignment configurations in multi-dimensional policies instead of having in MDP a single gate to flight assignment.

However, MMDP model gives only fixed time steps between decision epochs (iteration steps), that does not reveal the real evolution of gate assignment witch time is different from iteration step and always observable. Next paragraph gives an experiment with time dependence.

B. Time-Dependent Multi-Agent Model Experiment:

In this paragraph, it is conducted an experiment of Time-Dependent Multi-Agent MDP modeled earlier.

For simplification, every action possesses a single outcome. Hence actions and outcomes can be directly recognized ($a_i \leftrightarrow \mu_i$) and actions thought to be deterministic with regard to the discrete component of the state. This is expressed as:

$$\forall i \text{ Such that } a_i \text{ is feasible in state } s, L(\mu_i | s, t, a_i) = 1$$

It is used a real data from six flights of Hong Kong international airport as in Table IV; tree gates are dedicated to those flights.

A Gate conflict is detected between flights LH738/739 and SQ862/861 due to some disturbance.

Starting with a specific state of the system s_1 corresponding to the airport gate assignment:

$$s_1 = (\text{CA101/102, LH738/739, TG600/601})$$

TABLE IV. DATA FROM HONG KONG INTERNATIONAL AIRPORT [33]

Flight	Arrival	Departure	Route	Airline
CA101/102	11:25	12:45	Beijing-Hong Kong-Beijing	Air China
LH738/739	11:30	13:10	Frankfurt-Hong Kong-Frankfurt	Lufthansa
TG600/601	11:45	12:45	Bangkok-Hong Kong-Bangkok	Thai Airway
JL710/702	13:15	15:00	Osaka-Hong Kong-Osaka	Japan Airlines
BR869/870	14:25	15:30	Taipei-Hong Kong-Taipei	EVA Air
SQ862/861	14:20	16:00	Singapore-Hong Kong-Singapore	Singapore Airlines

Moreover, exploiting other possible actions is done to apply adapting assignment to arriving flights representing a change in gate configuration.

- $a_1 = (\text{BR869/870}, \text{JL710/702}, \text{SQ862/861})$
- $a_2 = (\text{BR869/870}, \text{SQ862/861}, \text{JL710/702})$
- $a_3 = (\text{JL710/702}, \text{SQ862/861}, \text{JL710/702})$

Fig. 8 below shows the state transition corresponding diagram.

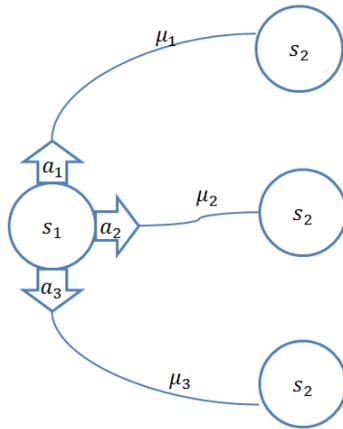


Fig. 8. State transition diagram.

Just for simplification, all outcomes have parameter $T_{\mu} = ABS$, so outcomes with durations are not considered. The probability density functions P_{μ} are the defined for every outcome see as example Fig. 9.

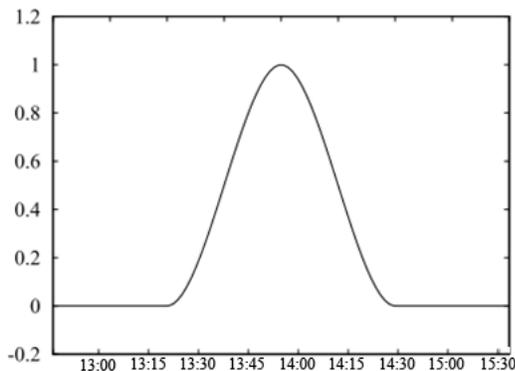


Fig. 9. Probability density functions of μ_2 .

This probability includes stochastic information related to action execution. Rewards are given in a way to score every action of assignment in the airport.

So, implementing the resolution algorithm, the value iteration algorithm gives an exact resolution [27]. The given solution consists of time-dependent policy choosing outcome μ_2 that avoid the disturbance situation. Then, the solution given by this approach is robust and handles flight delays. The fact of including the information about the possible disturbances improves more the GAP solution quality.

V. CONCLUSION AND PERSPECTIVE

In this work, A new approach has been formulated for the Gate Assignment Problem (GAP) powered by Time-dependent Multi-Agent Markov Decision Processes (TMMDP). This method aims to constitute a robust mechanism that will give a time valuated approach dealing with disturbances in every time sequence. The provided solution is all of the decisions at every time that could be performed at the time of the planning horizon of flights assignment. This kind of model takes into account real-time optimization because it assumes to have a solution at every time which manages disturbances.

Experimentations on this approach using a real sample data by simulation of the associated value iteration algorithm provides a best feasible solution that the deterministic model.

The aim behind this reflection is to offer to controllers at the airport a robust time valuated solution take in consideration possibilities of gate conflict, even if may take more time to resolution, it can manage well risks in gate assignment.

As perspective, this reflection about this type of model can be more extended to take into account as possible other real constraints of gate assignment.

REFERENCES

- [1] B. Pearce, "The state of air transport markets and the airline industry after the great recession," *Journal of Air Transport Management*, vol. 21, pp. 3–9, 2012.
- [2] A. Lim and F. Wang, "Robust airport gate assignment," in *Tools with Artificial Intelligence, 2005. ICTAI 05. 17th IEEE International Conference on*, Nov 2005, pp. 8 pp.–81.
- [3] C.-H. Cheng, S. C. Ho, and C.-L. Kwan, "The use of meta-heuristics for airport gate assignment," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12430–12437, Nov. 2012.
- [4] T. Obata, "Quadratic assignment problem: evaluation of exact and heuristic algorithms," 1979.
- [5] R. A. Howard, *Dynamic Programming and Markov Processes*. Cambridge, MA: MIT Press, 1960.
- [6] C. Boutilier, "Planning, learning and coordination in multiagent decision processes," in *In Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK96, 1996*, pp. 195–210.
- [7] J. Xu and G. Bailey, "The airport gate assignment problem: Mathematical model and a tabu search algorithm," in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 3 - Volume 3*, ser. HICSS '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 3032–.
- [8] U. Dorndorf, F. Jaehn, and E. Pesch, "Modelling robust flight-gate scheduling as a clique partitioning problem," *Transportation Science*, vol. 42, no. 3, pp. 292–301, Aug. 2008.

- [9] S. Yan and C.-M. Huo, "Optimization of multiple objective gate assignments," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 5, pp. 413–432, 2001.
- [10] A. Drexl and Y. Nikulin, "Multicriteria airport gate assignment and pareto simulated annealing," *IIE Transactions*, vol. 40, no. 4, pp. 385–397, 2008.
- [11] M. Hassounah and G. N. Steuart, "Demand for aircraft gates," *Transportation Research Record*, n 1423, pp.26-33, 1993.
- [12] S. Yan and C.-M. Chang, "A network model for gate assignment," *Journal of Advanced Transportation*, vol. 32, no. 2, pp. 176–189, 1998.
- [13] B. Mahajan and T. I. Matis, "Multi-commodity flow network model of the flight gate assignment problem," *Computers & Industrial Engineering*, vol. 63, no. 4, pp. 1135–1144, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360835212001799>
- [14] S. Yan and C.-H. Tang, "A heuristic approach for airport gate assignments for stochastic flight delays," *European Journal of Operational Research*, vol. 180, no. 2, pp. 547–567, 2007.
- [15] C.-M. Pinteá, P. Pop, C. Chira, and D. Dumitrescu, "A hybrid ant-based system for gate assignment problem," in *Hybrid Artificial Intelligence Systems*, ser. Lecture Notes in Computer Science, E. Corchado, A. Abraham, and W. Pedrycz, Eds. Springer Berlin Heidelberg, 2008, vol. 5271, pp. 273–280. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-87656-4_34
- [16] C. Yu, D. Zhang, and H. Lau, "Mip-based heuristics for solving robust gate assignment problems," *Computers & Industrial Engineering*, vol. 93, pp. 171–191, 2016.
- [17] W. Deng, H. Zhao, X. Yang, J. Xiong, M. Sun, and B. Li, "Study on an improved adaptive pso algorithm for solving multi-objective gate assignment," *Applied Soft Computing*, vol. 59, pp. 288–302, 2017.
- [18] S. Mokhtarimousavi, D. Talebi, and H. Asgari, "A Non-Dominated Sorting Genetic Algorithm Approach for Optimization of Multi-Objective Airport Gate Assignment Problem," *Transportation Research Record*, 2018, doi:10.1177/0361198118781386.
- [19] O. Aoun, M. Sarhani, and A. E. Afia, "Investigation of hidden markov model for the tuning of metaheuristics in airline scheduling problems," *IFAC-PapersOnLine*, vol. 49, no. 3, pp. 347–352, 2016, 14th IFAC Symposium on Control in Transportation SystemsCTS 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2405896316302543>
- [20] O. Aoun and A. El Afia, "Using markov decision processes to solve stochastic gate assignment problem," in *Logistics and Operations Management (GOL), 2014 International Conference on*, June 2014, pp. 42–47.
- [21] —, "Application of multi-agent markov decision processes to gate assignment problem," in *Information Science and Technology (CIST), 2014 Third IEEE International Colloquium in*, Oct 2014, pp. 196–201.
- [22] H. Yanco and L. A. Stein, "An adaptive communication protocol for cooperating mobile robots," in *Proceedings of the Second International Conference on Simulation of Adaptive Behavior*. MIT Press, 1993, pp. 478–485.
- [23] H. W. Kuhn and A. W. Tucker, "John von neumann's work in the theory of games and mathematical economics," *Bulletin of the American Mathematical Society*, vol. 64, pp. 100–122, 05 1958.
- [24] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *IN PROCEEDINGS OF THE ELEVENTH INTERNATIONAL CONFERENCE ON MACHINE LEARNING*. Morgan Kaufmann, 1994, pp. 157–163.
- [25] O. Aoun and A. El Afia, "A robust crew pairing based on multi-agent markov decision processes," in *Complex Systems (WCCS), 2014 Second World Conference on*, Nov 2014, pp. 762–768.
- [26] P. Stone and M. Veloso, "EnglishMultiagent systems: A survey from a machine learning perspective," *EnglishAutonomous Robots*, vol. 8, no. 3, pp. 345–383, 2000. [Online]. Available: <http://dx.doi.org/10.1023/A%3A1008942012299>
- [27] J. A. Boyan and M. L. Littman, "Exact solutions to time-dependent mdps," in *in Advances in Neural Information Processing Systems*. MIT Press, 2000, pp. 1026–1032.
- [28] S. Proper and P. Tadepalli, "Solving multiagent assignment markov decision processes," in *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, ser. AAMAS '09. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2009, pp. 681–688. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1558013.1558107>
- [29] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 3, pp. 387–434, Nov. 2005.
- [30] M. P. Wellman, K. Larson, M. Ford, and P. R. Wurman, "Path planning under time-dependent uncertainty," in *In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 1995, pp. 532–539.
- [31] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, ser. Wiley Series in Probability and Statistics. Wiley, 1994. [Online]. Available: <https://books.google.co.ma/books?id=tsiiQgAACAAJ>
- [32] G. S. Bastos, L. E. Souza, F. T. Ramos, and C. H. Ribeiro, "A single-dependent agent approach for stochastic time-dependent truck dispatching in open-pit mining," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE, 2011, pp. 1057–1062.
- [33] J. Wang, S. Hou, Y. Su, J. Du, and W. Wang, "Markov decision process based multi-agent system applied to aeroengine maintenance policy optimization," in *Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on*, vol. 3, Oct 2008, pp. 401–408.
- [34] J.-Y. Greff, L. Idoumghar, and R. Schott, "AnglaisApplication of markov decision processes to the frequency assignment problem," *AnglaisJournal on Applied Artificial Intelligence*, vol. 18, no. 8, pp. 761–773, 2004.