

Table des matières

Résumé	iii
Abstract	iv
Table des matières	v
Liste des figures	vii
Table des notations	viii
Remerciements	x
Introduction	1
1 Objets et définitions de base	7
1.1 Théorie des graphes	7
1.2 Processus aléatoires	9
1.3 Optimisation par Contraintes (CP)	11
2 Théorie des jeux et jeux de poursuite	14
2.1 Résolution de jeux finis	14
2.1.1 Résolution par relation d'ordre	14
2.1.2 Le point de vue de la théorie des jeux, le théorème de Kuhn	17
2.2 Les jeux stochastiques	18
2.2.1 Valeurs et stratégies optimales	22
3 Généralisation des jeux de poursuite stochastiques	27
3.1 Introduction	27
3.1.1 Concept de solutions dans les jeux de poursuite	29
3.2 Un jeu de poursuite abstrait à information parfaite	30
3.2.1 Stratégies	37
3.2.2 Conditions de victoire dans les jeux de policiers et voleurs	38
3.2.3 Résolution des jeux de policiers et voleurs	40
3.3 Complexité de calcul dans les jeux de poursuite	42
3.3.1 La complexité de calcul de la fonction w_n	42
3.3.2 Un résultat de difficulté pour les jeux de poursuite	44
3.4 Un résultat de stationnarité	45
3.5 Un modèle concret pour les jeux de policiers et voleurs	51
3.6 Exemples	56

3.7	Conclusion et travaux futurs	63
3.7.1	Jeux à information imparfaite	64
3.7.2	En conclusion	65
4	Applications des jeux de poursuite en recherche opérationnelle	67
4.1	Introduction	67
4.2	Théorie de la recherche et OSP	68
4.3	Le problème de l'OSP, ou le jeu du perdu	68
4.4	L'OSP en programmation par contraintes (CP)	72
4.4.1	Modélisation en CP	72
4.4.2	Le problème du perdu résolu par branch and bound	73
4.5	La borne DMEAN	74
4.6	Le jeu du policier et du voleur saoul et évusif	77
4.6.1	Le jeu du policier et du voleur saoul	77
4.6.2	La récurrence w_n	77
4.6.3	Le jeu du policier et voleur en processus de Markov décisionnel (MDP)	80
4.6.4	Quelques concepts de MDP	80
4.6.5	Retour au MDP du policier et voleur saoul	81
4.7	La borne COPWIN	83
4.8	Expérimentations et résultats	86
4.8.1	Résultats de la comparaison entre COPWIN et DMEAN	88
4.8.2	Résultats sur l'efficacité de COPWIN	91
4.9	Conclusion et travaux futurs	91
	Conclusion	94
	A Annexe	98
A.1	Le jeu de policiers et voleurs décrit comme un jeu stochastique simple . . .	98
	Bibliographie	102

Liste des figures

1.1	Le graphe de Petersen, son petit frère et un graphe générique.	8
1.2	Exemple de chemin (a), de cycle (b) et de graphe complet (c) à quatre sommets.	9
1.3	Chaîne de Markov à deux états A et B . Les probabilités de transition sont données sur les arêtes.	11
2.1	Trois graphes policiers-gagnants.	16
3.1	Exemple de deux mouvements des joueurs dans le jeu de policier et voleur saoul.	34
3.2	Illustration du troisième cas de la proposition 3.4.1	49
4.1	Le jeu du perdu sur une grille	69
4.2	Exemple de jeu de policier et voleur saoul.	78
4.3	Résumé des caractéristiques partagées et distinctives entre l'OSP et policier et voleur saoul	84
4.4	Certains des graphes utilisés pour les expérimentations.	87
4.5	Comparaison des valeurs objectifs de chaque borne	89
4.6	Nombre de coupes de branches par borne	90
4.7	Résultats sur différentes instances du problème OSP ; les valeurs en gras sont meilleurs.	92
4.8	Valeur objectif obtenue en fonction du temps	93
4.9	Comparaison des valeurs objectifs de chaque borne avec l'heuristique TD	93

Table des notations

$(x_i)_{i \in \{1, \dots, n\}}$	Suite d'éléments x indicés de 1 à n
V	Ensemble des sommets d'un graphe
E	Ensemble des arêtes d'un graphe
$[x, y]$	Arête entre les sommets x, y d'un graphe
$N[c]$	Le voisinage fermé de c dans un graphe, incluant c
$N(c)$	Le voisinage ouvert de c dans un graphe, excluant c
\mathcal{P}	L'ensemble des chemins de longueur finie d'un graphe
\mathcal{P}_u	L'ensemble des chemins de longueur finie d'un graphe débutant en un sommet u
Δ	Le degré maximal des sommets d'un graphe
\mathcal{G}	Jeu de poursuite général
$T_{\sigma_{\text{cop}}, \sigma_{\text{rob}}}$	Temps de capture du voleur sous les stratégies $\sigma_{\text{cop}}, \sigma_{\text{rob}}$
p_N^*	Probabilité que le temps de capture soit inférieur ou égal à N
\preceq_n	La relation d'ordre de Nowakowski et Winkler ou de Bonato et MacGillivray
w_n	Relation de récurrence donnant la probabilité de capture en n tours
$S_{\text{cop}}, S_{\text{rob}}, S_o$	Les ensembles d'états policiers, voleurs et autres
$A_{\text{cop}}, A_{\text{rob}}$	Les ensembles d'actions des policiers et des voleurs
$T_{\text{cop}}, T_{\text{rob}}$	Les fonctions de transition des policiers et des voleurs
$c(p, n)$ -gagnant	Un jeu de policiers et voleurs dans lequel $w_n(i) \geq p$
$c_p^n(\mathcal{G})$	Le nombre minimal de policiers nécessaire à ce qu'un jeu \mathcal{G} soit $c(p, n)$ -gagnant
Dist_B	L'ensemble des distributions sur un ensemble B
\mathcal{U}_B	La distribution uniforme sur un ensemble B
δ_s	La distribution de Dirac sur un élément s
$\text{pod}(v)$	Probabilité de détection du perdu sur le sommet v

$\text{poc}_{\pi_t}(v)$	Probabilité de présence du perdu sur le sommet v conditionnellement au chemin π_t du chercheur
cos_{π_t}	Probabilité de capture du perdu conditionnellement au chemin π_t du chercheur
$\text{cos}_{\pi_T}^*$	Probabilité de capture maximale du perdu en T incréments de temps, où π_T^* est le chemin maximisant cette probabilité parmi tous les chemins de longueur T

Remerciements

Voici enfin arrivé l'aboutissement d'un périple. En cette fin de mémoire, j'aimerais remercier tous ceux qui ont de près ou de loin, par leurs encouragements, soutiens et enseignements, contribué à la réalisation de cet ouvrage. La qualité de ce travail n'aurait pu culminer en cette valeur sans l'apport formidable de mes directeurs de recherche, les professeurs François Laviolette et Josée Desharnais. Je leur communique mes remerciements de m'avoir partagé un enseignement qui me fut et qui continuera à m'être fort utile. D'un même trait, je remercie les évaluateurs qui ont accordé de leur temps pour lire ce document et suggérer des commentaires constructifs pour en accroître la richesse. Il m'est aussi vital de saluer Michael Morin, Ph. D., sans qui une des contributions majeures de ce mémoire, le chapitre 4, n'aurait pu voir le jour. Je remercie encore les professeurs avec qui j'ai eu la chance de travailler au cours de ces années de maîtrise dans divers projets de recherche : Luc Lamontagne, Ismaïl Biskri, Josée St-Pierre, Claude-Guy Quimper, Raphaël Khoury, Yezekael Hayel et Dominique Quadri. L'expérience que j'ai pu acquérir aux côtés de ces chercheurs m'est précieuse et m'a grandement permis d'affiner ma vision de la recherche.

Plus personnellement, ces années de recherche n'auraient pu être complétées sans le soutien de ma famille. Je remercie ma mère qui m'a notamment partagé sa connaissance des subtilités du monde académique, mon père qui m'a offert son soutien avec ses lumineuses perles de sagesse, dont la très fameuse « Ça se fait par du monde. », et mon frère pour son apport, disons, social. Enfin, cette recherche aurait été plus que maussade sans la présence de mes amis proches, que je salue : Maxime et Maxime, Vincent, Annick et Anne. Évidemment, la liste des personnes remerciées n'est pas exhaustive, je tiens aussi à exprimer mes remerciements à tous les gens qui ont été oubliés.

Introduction

Historique

Un formalisme généralement simple, mais efficace, pour décrire les interactions entre des agents dans des contextes parfois très complexes est celui de la théorie des jeux. Développé il y a de cela près d'une centaine d'années avec de premiers résultats dus à Zermelo¹, suivis par une découverte plus formelle et globale due à Newmann et Morgenstern pour finalement aboutir au théorème novateur de Nash², la théorie des jeux est maintenant incontournable dans des domaines aussi éloignés que l'intelligence artificielle et l'économie. Notamment, la théorie des jeux a eu un impact profond dans une branche importante de la recherche opérationnelle qu'est la théorie de la recherche qui étudie des questions pratiques telles que la rescousse d'individus perdus en mer. On parle généralement de *jeux* pour signifier un contexte précis établi avec des *règles* régissant les interactions permises entre des *agents* ainsi que les objectifs que ceux-ci cherchent à accomplir.

Un *jeu de poursuite* est un problème inspiré par la théorie des jeux visant à retrouver un individu se promenant dans un environnement précis. Déjà, lorsque la première formulation d'un jeu de policier et voleur fut établie par Nowakowski et Winkler [39] en 1983, et indépendamment par Quilliot [46], près de vingt années s'étaient écoulées depuis la première formulation connue d'un jeu de poursuite due à Isaacs avec son jeu *homicidal chauffeur* [23]³. Dans ce jeu, un piéton (l'évadé) est poursuivi par un chauffeur (le poursuivant) avec l'intention de l'écraser. Isaacs cherchait à comprendre sous quelles conditions le chauffeur pourrait atteindre

1. Zermelo est considéré comme étant le premier à avoir formulé des résultats de théorie des jeux en démontrant qu'aux Échecs soit les blancs peuvent être assurés de gagner, soit les noirs peuvent être assurés de gagner, soit les deux joueurs sont certains de faire une partie nulle. Ce résultat est démontré dans un article de 1913 en allemand, mais il semble malheureusement qu'aucune traduction française ou anglaise n'existe. Certains lui créditent des résultats plus généraux, cependant. On renvoie à l'article de Ulrich et Walker [55] pour plus de détails sur cette partie de l'histoire de la théorie des jeux.

2. L'histoire de la théorie des jeux ainsi que les résultats importants de Newman, Morgenstern et Nash peuvent se trouver dans un bon livre de théorie des jeux, tel que ceux de Tadelis ou Shoham et Leyton-Brown [53, 51]. Brièvement, Newman et Morgenstern ont apporté la notion de fonction d'utilité qui permet de définir les objectifs d'un joueur ainsi que l'analyse de jeux très simples à deux joueurs. Nash a donné son nom à *l'équilibre de Nash* qui décrit une situation dans laquelle chaque joueur joue de manière optimale dans des jeux quelconques avec un nombre arbitraire de joueurs.

3. La première formulation de ce jeu fut donnée dans un texte soumis à la corporation RAND en 1951, mais Isaacs le représente dans son livre de 1965, *Differential Games* à côté de plusieurs autres jeux de poursuite.

son objectif. Isaacs a continué à travailler dans les jeux de poursuite pour ensuite formuler un autre jeu assez connu, celui du *monstre et de la princesse* [23] dans lequel un monstre cherche à capturer une princesse dans une chambre complètement sombre. Quoique les jeux de Isaacs se décrivent en temps continu alors que celui du policier et voleur se déroule en temps discret, on peut tracer un parallèle entre les deux via les jeux intermédiaires qui furent étudiés durant ces vingt années. En effet, depuis Isaacs, plusieurs articles et livres ont été écrits sur les jeux de poursuite. Évidemment, les méthodes de résolution de ces deux exemples diffèrent beaucoup : Isaacs a développé les équations différentielles éponymes qui permettent de résoudre le jeu du monstre et de la princesse alors que Nowakowski et Winkler ont élaboré une relation de récurrence⁴.

Dans ce mémoire, nous sommes intéressés par les jeux à temps discret et se déroulant sur des structures discrètes, principalement des graphes, dans lesquels les équations différentielles de Isaacs ne sont pas définies. En 1976, selon la recension de littérature de Fomin et Thilikos [17], fut formulé le premier jeu de poursuite sur graphe. Celui-ci se déroulait toutefois en temps continu. Parson [43] formula ce premier jeu pour décrire le mouvement de spéléologues dans une caverne. Certains jeux différentiels (jeux se décrivant avec des équations différentielles) furent étudiés sous la restriction que l’environnement soit un graphe alors que ces jeux sont typiquement conçus pour des environnements continus. Petrov [44] a indépendamment redécouvert le problème ainsi que certains résultats de Parsons plus tard. Golovach a démontré en 1989 [22, 21] que les modèles de Parsons et Petrov étaient équivalents à un jeu de poursuite à temps discret sur un graphe. Dès lors fut complétée la transition vers les jeux purement discrets (en temps et en espace). L’évadé, dans le jeu de Golovach, est invisible, ce qui complique la tâche des poursuivants. Ce jeu fait encore aujourd’hui partie d’un domaine actif de recherche portant le nom des jeux de *fouille de graphes* dans lesquels un évadé complètement invisible est recherché sur un graphe.

Jeux de policiers et voleurs

En 1983, pour mener jusqu’au bout la discrétisation et la simplification des jeux de poursuite, Nowakowski et Winkler [39] et, indépendamment, Quilliot [46], ont défini puis résolu, c’est-à-dire ont déterminé la stratégie optimale du policier⁵, un premier jeu de policier et voleur. Celui-ci se joue encore une fois à deux joueurs, l’un poursuivant l’autre à information parfaite⁶, sur un graphe et à tour de rôle. Depuis, plusieurs articles ont été publiés concernant ce type de jeux qui, chacun, font varier un paramètre du jeu (vitesse des joueurs, type de graphe, etc.).

4. On reviendra plus tard, dans le chapitre 1, sur la notion de *graphes*, aussi appelés *réseaux*. Pour l’instant, il est simplement important de penser à un graphe comme un ensemble des points reliés par des droites. On pense par exemple à un réseau d’ordinateurs, une carte routière, un réseau social, etc.

5. Dans les jeux de policiers et voleurs, on dit qu’un jeu est résolu soit lorsque l’on peut déterminer avec précision quel joueur va gagner, soit lorsque la probabilité de capture des voleurs par les policiers est calculée. L’expression est tirée de la théorie des jeux où l’on dit qu’un jeu est résolu lorsque l’ensemble des gains des joueurs, en supposant que ceux-ci jouent optimalement, a été calculé.

Le livre de Bonato et Nowakowski, [8], recensant l'histoire de ce type de jeux et son intérêt en mathématiques et en informatique est d'ailleurs une très bonne référence sur le sujet. Il est coutume en mathématiques de vouloir simplifier un problème ardu pour plus aisément le résoudre. C'est possiblement ce qui a mené Nowakowski et Winkler, et Quilliot, à décrire un jeu dans lequel l'évadé est complètement visible pour contribuer à résoudre d'autres jeux de poursuite comme celui de Golovach.

Ces jeux, de par leurs formulations simples, suscitent aussi beaucoup d'intérêt à différents niveaux d'abstraction. Certains y voient des modèles de caractérisation de paramètres importants de graphes tels que la largeur d'arbre ou de chemin (on renvoie à l'article de Fomin et Thilikos [17] pour plus d'explications sur les liens entre les jeux de poursuite et les paramètres de graphes ainsi qu'au livre de Diestel [16] pour plus d'information sur ces paramètres). C'est d'ailleurs une des raisons pour laquelle on classe les graphes selon les solutions qu'il produisent sous certains jeux : l'ensemble des graphes sur lesquels k policiers peuvent capturer un voleur sous telles règles, l'ensemble des graphes sur lesquels l voleurs peuvent se sauver sous telles règles, etc. D'autres utilisent certaines variantes pour modéliser des intrusions de virus dans des systèmes informatiques [2]. Finalement, d'autres encore, tels que dans le présent travail, utilisent la proximité de ces jeux avec des applications connues en théorie de la recherche pour aider à résoudre des problèmes de fouilles de graphes qui sont difficiles au niveau computationnel (NP-difficiles [58], par exemple).

L'intérêt pour les jeux de policiers et voleurs est donc bien présent, ce qui justifie l'étude des différentes variantes formulées dans les dernières années.

Le point de vue de la théorie des jeux

Un aspect souvent négligé des jeux de policiers et voleurs est leur proximité avec la théorie des jeux en général, aspect qui pourrait être utile pour résoudre des problèmes complexes apparaissant dans ces jeux tels que le problème du traitement des jeux à information partielle. En effet, malgré que les jeux de policiers et voleurs soient en général traité comme des problèmes distincts naissant naturellement de la théorie des graphes, et donc n'utilisant en général pas le même langage que la théorie des jeux, on peut aisément trouver de grandes similarités entre les deux formalismes.

On peut faire de beaux rapprochements entre la manière dont sont résolus typiquement les jeux de policiers et voleurs et les méthodes classiques de la théorie des jeux. En effet, on peut remarquer qu'une méthode assez classique pour déterminer la stratégie optimale des policiers peut être réécrite comme une variante d'une méthode connue de la théorie des jeux.

6. Au chapitre 1 sont présentées plus en détail les définitions propres à la théorie des jeux telles que la notion d'information parfaite.

D'un autre côté, un des aspects les plus importants de ce mémoire est l'intégration d'éléments aléatoires dans les jeux de policiers et voleurs. En effet, ceux-ci sont typiquement purement déterministes, c'est-à-dire ils ne supposent aucun aspect stochastique. Cependant, une partie de la recherche sur les jeux de policiers et voleurs se dirige depuis quelques années vers des jeux dans lesquels, par exemple, un voleur se déplace aléatoirement [25, 28]. De tels jeux ne peuvent être résolus à l'aide de la même méthode de résolution utilisée dans les jeux de policiers et voleurs déterministes. Il est donc nécessaire de développer de nouveaux outils pour inclure ces nouvelles éventualités. Une certaine branche de la théorie des jeux, dite des *jeux stochastiques*, présente un cadre adéquat et élégant pour intégrer ces éléments aléatoires. Il semble donc naturel d'exposer cette théorie dans ce mémoire. On verra au chapitre 3 que celle-ci n'a cependant pas été appliquée directement dans nos travaux, elle est donc présentée simplement à des fins d'exposition.

En finir avec les jeux de policiers et voleurs

Depuis les trente dernières années, une grande variété de jeux de policiers et voleurs a été énoncée et résolue. Comme on peut le remarquer dans les revues de littérature sur le sujet telles que le livre de Bonato et Nowakowski [8], la quantité de variantes du jeu original de Nowakowski et Winkler est très impressionnante. Cependant, chaque variante a historiquement été étudiée en vase clos et on peut aisément justifier le désir de vouloir unifier les jeux de policiers et voleurs sous un même modèle. D'autant plus qu'en 2012, après quelque trente années de recherche, Clarke et MacGillivray [12] ont résolu le jeu à k policiers, essentiellement une des variantes les plus importantes de ce type de jeux, avec les mêmes outils que ceux utilisés pour résoudre le jeu à un seul policier. On peut donc flairer en lisant l'article de Clarke et MacGillivray qu'il devrait exister une manière de réunir tous les jeux de policiers et voleurs sous un même modèle et les résoudre une fois pour toutes.

Très récemment, Bonato et MacGillivray décrivaient un modèle en affirmant qu'il généralisait tous les jeux de policiers et voleurs [6]. À l'aide d'un formalisme très simple, ils y résolvaient l'ensemble des jeux de policiers et voleurs en généralisant la méthode de Nowakowski et Winkler. Un aspect n'était pas couvert cependant : depuis quelques années est étudié dans la littérature [28, 25] un jeu de policier et voleur, dit du policier et voleur saoul, dans lequel le voleur se déplace de manière aléatoire et donc sans réagir optimalement aux actions du policier. Ce jeu, résolu dans notre article de CP 2015 [52], n'entrait pas dans le cadre de Bonato puisque celui-ci n'y introduisait que des jeux totalement déterministes tels qu'ils furent toujours conçus avant qu'advienne le jeu de policier et voleur saoul.

Ainsi, la contribution principale de ce mémoire, au chapitre 3, est une généralisation de l'article de Bonato et MacGillivray à des contextes plus généraux où différents événements aléatoires pourraient se produire. Non seulement les voleurs sont-ils libres de bouger aléatoirement, mais

après chaque action entreprise par les joueurs peut survenir un événement aléatoire venant, par exemple, contrecarrer les actions de ceux-ci.

Le chapitre 3 répond donc à la problématique suivante :

Suite à la présentation du modèle général de jeux de poursuite de Bonato et MacGillivray, il appert que les jeux intégrant des composantes stochastiques ne sont pas pris en compte. Est-il possible, et si oui comment, de concevoir un nouveau cadre général incluant le premier ainsi que les aspects aléatoires n'ayant pas été considérés ?

Le problème de l'information

Un aspect fondamental des jeux de poursuite, et de tout jeu en général, est la question de la quantité d'information disponible aux protagonistes. Plus particulièrement, on s'intéresse généralement à la quantité d'information que possèdent les poursuivants. Aux deux extrêmes d'un large spectre on retrouve les jeux à information nulle et les jeux à information parfaite : d'une part lorsque l'information est nulle les poursuivants n'ont aucune information sur la position des évadés, on parle par exemple des jeux d'*Optimal Search Path* ou de fouille de graphe, d'autre part dans un jeu à information parfaite les poursuivants ont en tout temps connaissance de la position des évadés tels que dans les jeux de policiers et voleurs. Entre ces deux bornes se compte une vaste quantité de jeux différents tels que les jeux avec témoins, caméras, radars, etc. qui diffusent une quantité limitée et ponctuelle d'information sur la position des évadés.

Naturellement, dans un même jeu plus la quantité d'information disponible aux poursuivants augmente, plus élevée est la probabilité que ceux-ci en viennent à capturer leur adversaire. Ceci est en particulier vrai pour les jeux de policiers et voleurs : on peut montrer qu'un problème de recherche opérationnelle dénommé le *Optimal Search Path* (OSP) dans lequel un individu cherche un évadé lui étant invisible sur un graphe se traduit en un jeu de policier et voleur lorsque la position de l'évadé est révélée au chercheur. C'est cette association qui a été exploitée dans l'article présenté à la conférence « 21st International Conference on Principles and Practice of Constraint Programming (CP 2015) » [52].

On peut formellement énoncer la question de recherche auquel le chapitre 4 prétend répondre de la manière suivante :

Compte tenu de la proximité entre le problème de recherche dit de l'OSP et des jeux de policiers et voleurs, serait-il possible, et approprié, de définir formellement une borne supérieure admissible sur la fonction objectif de ce premier problème à l'aide d'un jeu de policier et voleur ?

Cette problématique est plus amplement développée au chapitre 4 dans lequel est retranscrit

l'article qui a été accepté et présenté à la conférence « 21st International Conference on the Principles and Practice of Constraint Programming » [52], à Cork en Irlande. Celui-ci expose comment résoudre d'une manière relativement simple un jeu de poursuite avec évadé invisible en utilisant sa caractérisation sous la forme d'un jeu de policier et voleur. On y montre comment accélérer la résolution du problème de l'OSP à l'aide d'une relaxation de contraintes vers un jeu de policier et voleur.

Organisation de ce mémoire

En premier lieu, le chapitre 1 présente des définitions de base d'objets mathématiques qui seront utilisés dans ce mémoire. Plus précisément, on y présente la théorie des graphes, la théorie des processus aléatoires ainsi que la programmation par contraintes.

Ensuite sont présentées au chapitre 2 les méthodes généralement utilisées pour trouver la stratégie optimale des policiers dans les jeux de policiers et voleurs ou, autrement dit, pour *résoudre* ces jeux. Ce chapitre présente en quelque sorte les outils de base utilisés dans les jeux de policiers et voleurs, outils qui se retrouvent dans les chapitres qui suivent. Ce chapitre présente aussi comment la méthode de résolution des jeux de policiers et voleurs se compare à la méthode de résolution de jeux généraux de la théorie des jeux. Finalement, une branche de la théorie des jeux, la théorie des *jeux stochastiques*, est exposée de manière plus formelle.

Au chapitre 3 est présentée une des contributions majeures de ce mémoire. Celle-ci fera d'ailleurs l'objet d'un article de journal à être soumis. Dans ce chapitre sont développés un modèle générique de jeu de policiers et voleurs ainsi qu'une méthode récursive permettant de résoudre une infinité de jeux de policiers et voleurs différents. Ce chapitre, quoique s'inspirant d'un article déjà soumis de Bonato et MacGillivray, représente un progrès important dans la théorie des jeux de policiers et voleurs puisqu'il permet l'étude d'une infinité de jeux différents à l'aide d'un seul modèle en plus d'étendre la méthode de résolution classique des jeux de policiers et voleurs au cas stochastique.

Le chapitre 4 présente notre contribution à l'avancement de la recherche sur un problème connu de fouille de graphes, le problème OSP. Cette contribution est originale dans sa manière d'aborder le problème de l'OSP. À la lecture de l'état de l'art sur le sujet, il semble que la relaxation du problème en un jeu à information parfaite n'ait jamais été formellement établie. Cette publication ouvre la voie à l'application de cette méthode à d'autres problèmes du type OSP dont celui avec plusieurs chercheurs.

Les deux chapitres principaux de ce mémoire sont donc les chapitres 3 et 4. Comme ceux-ci traitent de deux problématiques distinctes, les revues de littérature concernant ces deux sujets seront directement incluses dans ceux-ci.

Chapitre 1

Objets et définitions de base

Certaines définitions d'objets mathématiques de base sont présentées dans ce chapitre, ceci dans le but de faciliter la lecture du reste du manuscrit. Les définitions présentées dans ce chapitre se divisent en trois grands thèmes : la théorie des graphes, les processus aléatoires et la programmation par contraintes. Il n'est nullement prétendu dans ce mémoire que ces définitions couvrent l'entièreté de leurs domaines respectifs et nous encourageons le lecteur intéressé à consulter les références recommandées pour obtenir de plus amples informations. Certaines sections utilisent des notations propres à leurs domaines respectifs, celles-ci sont résumées dans la table de notation située en page viii.

1.1 Théorie des graphes

Les définitions présentées dans cette section peuvent toutes être retrouvées dans un livre de théorie des graphes. Les références classiques dans ce domaine sont les ouvrages de Diestel [16], Bondy et Murty [9] et Bollobás [5].

Un graphe $G = (V, E)$ s'écrit comme un ensemble dénombrable V d'éléments appelés *sommets*, ainsi qu'une relation symétrique $E \subseteq V \times V$ dont les éléments sont nommés *arêtes*. On écrit $[u, v] \in E$ l'arête formée par les sommets $u, v \in V$. On représente visuellement les sommets par des points et les arêtes par des liens entre ceux-ci. Sauf indication contraire, on ne dessine qu'une seule arête par paire de sommets, pour simplifier. On écrit souvent $V(G)$ et $E(G)$ pour préciser de quel graphe on regarde l'ensemble de sommets ou d'arêtes, surtout lorsqu'il y a possibilité d'ambiguïté.

Trois graphes sont présentés à la figure 1.1. Le premier, un exemple de graphe célèbre présenté à la figure 1.1(a), est le graphe de *Petersen*. Le second, à la figure 1.1(b) est parfois appelé *le petit frère de Petersen* pour des raisons apparentes.

On écrit généralement n et m pour les cardinalités de l'ensemble de sommets et celle de l'ensemble d'arêtes. On parle de sommets *adjacents* lorsque deux sommets sont liés par une arête.

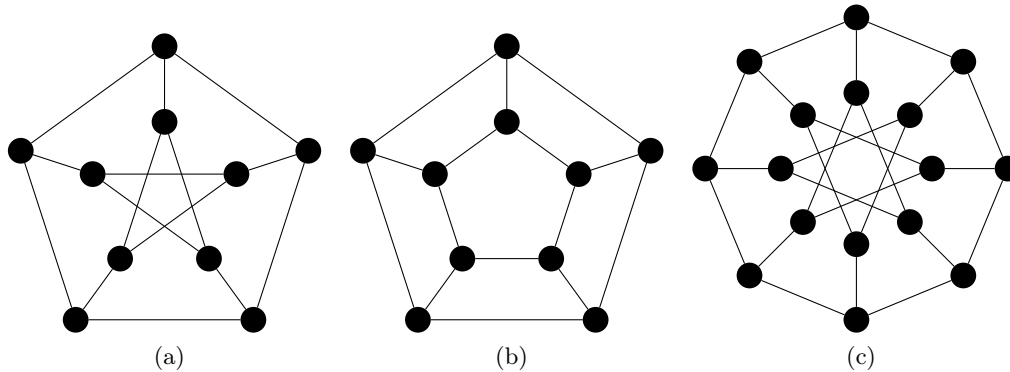


FIGURE 1.1 – Le graphe de Petersen, son petit frère et un graphe générique.

La dénombrabilité de V permet de traiter G comme une structure discrète, par opposition à un ensemble non dénombrable, ou continu, tel que \mathbb{R} . Un graphe peut être fini ou infini selon la cardinalité de son ensemble de sommets. On suppose toujours qu'il n'existe qu'une seule arête entre deux sommets (autrement on parlerait de multigraphe).

Un graphe H est un *sous-graphe* de G , noté $H \subseteq G$, si $V(H) \subseteq V(G)$ et $E(H) \subseteq E(G)$. Autrement dit, H est formé à l'aide d'une partie des sommets du graphe G et de certaines de ses arêtes. On voit par exemple que le graphe présenté à la figure 1.2(a) est un sous-graphe du graphe présenté à la figure 1.2(b), lui-même sous-graphe de celui illustré à la figure 1.2(c).

Plusieurs graphes simples ont leurs noms propres tels que le *chemin*, le *cycle* ainsi que le *graphe complet* (aussi appelé clique). Un chemin dans un graphe $G = (V, E)$ est formé par une suite $(v_i)_{i \in \{1, \dots, n\}}$, $n \geq 1$ de sommets distincts tels que chaque sommet est lié par une arête à son suivant, c'est-à-dire $[v_i, v_{i+1}] \in E$ pour tout $i \in \{1, \dots, n-1\}$. On définit la longueur d'un chemin par le nombre d'arêtes le composant. On note \mathcal{P} l'ensemble des chemins de longueur finie d'un graphe et \mathcal{P}_u l'ensemble des chemins de longueur finie débutant en u . Un cycle est un chemin contenant l'arête $[v_n, v_1]$. Le graphe complet est un graphe dont tous les sommets sont adjacents. On dénote par P_n , C_n , K_n respectivement le chemin (P venant du terme anglais « *path* »), le cycle et la clique à n sommets. Les figures 1.2(a), 1.2(b), 1.2(c) présentent respectivement les graphes P_4 , C_4 et K_4

On dit qu'un graphe G est *connexe* lorsque entre chaque paire (u, v) de sommets de G il existe un sous-graphe de G formant un chemin P débutant en u et se terminant en v ¹. Autrement dit, un graphe est connexe s'il est possible de le parcourir en entier en partant de n'importe quelle position et en suivant seulement les arêtes.

On écrit $N_G[v]$ pour dénoter le voisinage fermé d'un sommet v sur un graphe G (lorsqu'il n'y a aucune ambiguïté avec le contexte, on simplifie à $N[v]$), soit l'ensemble des sommets adjacents

1. Plus généralement, on parle de k -connexité pour un graphe ayant cette propriété avec « il existe un chemin » remplacé par « il existe k chemins distincts »

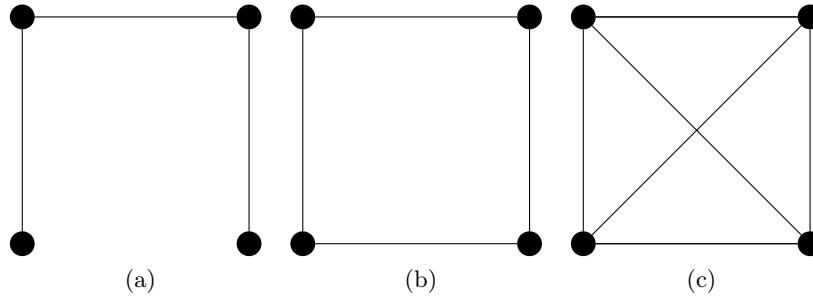


FIGURE 1.2 – Exemple de chemin (a), de cycle (b) et de graphe complet (c) à quatre sommets.

à v incluant lui-même

$$N[v] := \{u \in V : [u, v] \in E\} \cup \{v\}.$$

Similairement, le voisinage ouvert $N(v)$ d'un sommet v est l'ensemble des sommets voisins d'un sommet excluant lui-même, $N(v) = N[v] \setminus \{v\}$. Le degré d'un sommet $u \in V$, noté $\deg(u)$, est simplement la cardinalité du voisinage ouvert de u . On note aussi $\Delta = \max_{v \in V} \deg(v)$ le degré maximal d'un graphe, soit le degré le plus élevé parmi l'ensemble des sommets du graphe.

Finalement, on dit qu'un graphe est dirigé si la relation E n'est pas symétrique. Autrement dit, si $[u, v] \in E \not\Rightarrow [v, u] \in E$. Les notions décrites jusqu'à présent restent bien définies. On illustre chaque arête $[u, v] \in E$ à l'aide d'une flèche pour différencier entre les orientations des arêtes.

1.2 Processus aléatoires

Cette section présente quelques notions de base de processus aléatoires qui peuvent sans difficulté être retrouvées et approfondies dans un bon livre de probabilités avancées tel que ceux de Ross [48, 49]. En particulier, on définit les deux notions principales utilisées dans ce mémoire, soit la *chaîne de Markov* et son cas particulier, la *marche aléatoire*.

On utilise dans ce mémoire le terme général de *processus aléatoire* (ou processus stochastique) pour décrire une suite de variables aléatoires $(X_i)_{i \in \mathbb{N}}$ dont l'hypothèse i.i.d. (indépendantes et identiquement distribuées) n'est pas supposée. Ainsi, un processus aléatoire peut décrire une suite de variables dépendantes. Ces processus sont généralement utilisés pour décrire l'évolution, par exemple dans le temps ou l'espace, d'une variable aléatoire X . Notamment, le cours de la bourse peut se décrire à l'aide d'un processus aléatoire : si la valeur de la bourse (disons du Dow Jones) au temps t est donnée par X_t , alors chaque paire de variables X_t, X_{t+1} doit donc être dépendante. En effet, la valeur de la bourse au temps t doit influencer sur la valeur de celle-ci au temps $t+1$. Formellement, si $\mathbb{P}[X_i = x_i]$ est la probabilité qu'une variable aléatoire X_i prenne la valeur x_i , alors un processus aléatoire vise généralement à décrire la

probabilité conditionnelle $\mathbb{P}[X_t = x_t \mid X_{t-1} = x_{t-1}, \dots, X_0 = x_0]$, soit la probabilité qu'une variable prenne une certaine valeur étant donné l'ensemble des informations connues.

On rappelle que la probabilité conditionnelle $\mathbb{P}[A \mid B]$ décrit la probabilité que l'événement A se produise *conditionnellement* au fait que l'événement B s'est produit. Cette probabilité est définie par $\mathbb{P}[A \mid B] := \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$ (lorsque $\mathbb{P}[B] > 0$). Autrement dit, la probabilité conditionnelle est fonction de la probabilité que deux événements se produisent ($\mathbb{P}[A \cap B]$) et de la probabilité de réalisation d'un de ceux-ci ($\mathbb{P}[B]$). Plus de détails peuvent être trouvés dans tout bon texte d'introduction à la théorie des probabilités tel que [49].

Une chaîne de Markov est un processus aléatoire $(X_i)_{i \in \mathbb{N}}$ défini sur un espace d'états dénombrable S , c'est-à-dire chaque variable $X \in (X_i)_{i \in \mathbb{N}}$ est pigée dans S , et possédant les propriétés de Markov et d'homogénéité dans le temps.

$$\begin{aligned} \mathbb{P}[X_i = x_i \mid X_{i-1} = x_{i-1}, \dots, X_0 = x_0] &= \mathbb{P}[X_i = x_i \mid X_{i-1} = x_{i-1}] \quad (\text{propriété de Markov}) \\ \mathbb{P}[X_i = x \mid X_{i-1} = y] &= \mathbb{P}[X_{i+t} = x \mid X_{i-1+t} = y] \quad (\text{homogénéité dans le temps}) \end{aligned}$$

Ainsi, selon la propriété de Markov, la probabilité que la variable X_i prenne une certaine valeur x_i ne dépend pas de l'ensemble des valeurs prises par les variables précédentes x_{i-1}, \dots, x_0 , mais seulement de la valeur précédente x_{i-1} . D'un autre côté, l'homogénéité dans le temps de la chaîne de Markov nous assure que peu importe à quel moment une variable prend la valeur y , la probabilité que la variable suivante vaille x est invariante.

Ces deux propriétés sont illustrées sur l'exemple classique de la chaîne de Markov à deux états à la figure (1.3). On peut généralement représenter une chaîne de Markov comme un automate avec des transitions probabilistes. Ainsi, chaque état de S , l'ensemble d'états, est représenté par un sommet. Ici, S ne contient que deux états, A et B . Les transitions entre les états sont réalisées de manière probabiliste selon la chaîne de Markov. Imaginons donc qu'on ait un processus $(X_i)_{i \in \mathbb{N}}$ sur l'espace d'états $S = \{A, B\}$. Par la propriété de Markov on peut écrire la probabilité que la variable X_i prenne une certaine valeur seulement en fonction de son état précédent, soit $\mathbb{P}[X_i = B \mid X_{i-1} = A]$ ou $\mathbb{P}[X_i = A \mid X_{i-1} = A]$. L'homogénéité dans le temps de la chaîne de Markov permet de laisser tomber les indices temporels et d'écrire simplement $\mathbb{P}[X_i = B \mid X_{i-1} = A] = \mathbb{P}[B \mid A] = p$ puisque peu importe à quel moment le processus atteint l'état A , la probabilité de rencontrer ensuite l'état B sera toujours la même. Les boucles autour des états A et B complètent les probabilités de transiter aux autres états puisqu'une probabilité doit toujours sommer à 1. On a donc $\sum_{s \in \{A, B\}} \mathbb{P}[s \mid A] = p + (1 - p) = 1$.

Ensemble, les deux propriétés de la chaîne de Markov nous permettent de décrire une chaîne de Markov simplement à l'aide d'une matrice M telle que pour chaque élément (a, b) de M , on ait $M(a, b) = \mathbb{P}[b \mid a]$. Pour bien définir la chaîne de Markov, on ajoute à M une distribution initiale, parfois notée π , donnant la distribution de la variable X_0 , dans ce cas-ci

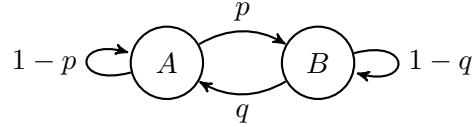


FIGURE 1.3 – Chaîne de Markov à deux états A et B . Les probabilités de transition sont données sur les arêtes.

$\pi = (\mathbb{P}[X_0 = A], \mathbb{P}[X_0 = B])$. Ainsi, dans le cas de la chaîne de Markov à deux états on a

$$M = \begin{matrix} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{matrix}.$$

La distribution initiale peut être donnée, par exemple, par

$$\pi = \left(\frac{1}{2}, \frac{1}{2} \right),$$

signifiant qu'avec probabilité $\frac{1}{2}$ la chaîne débute en l'état A ou en l'état B . Au chapitre 3 sera utilisée une propriété étonnante des chaînes de Markov, soit que la probabilité qu'en k incréments de temps un processus transite d'un état a à un état b est encodée dans la k -ième puissance de sa matrice M . Autrement dit,

$$M^k(a, b) = \mathbb{P}[X_{i+k} = b \mid X_i = a].$$

Une *marche aléatoire sur un graphe* G [5] est un cas particulier d'une chaîne de Markov lorsque l'espace d'états constitue les sommets d'un graphe et où ne sont permises que les transitions suivant des arêtes. Autrement dit, une marche aléatoire est une chaîne de Markov telle que chaque variable X_i est pigée dans l'ensemble des sommets du graphe $X_i \in V(G)$ et se retrouve dans le voisinage de la précédente, c'est-à-dire $X_{i+1} \in N[X_i]$. Ainsi, on peut imaginer l'évolution du processus comme celle d'un marcheur se promenant de sommets en sommets en choisissant toujours aléatoirement sa prochaine position parmi son ensemble de voisins. Une marche aléatoire, tout comme une chaîne de Markov générale, peut aussi être représentée simplement par sa matrice de transition M . M est ici une matrice $|V(G)| \times |V(G)|$. Chaque élément $(i, j) \in M$ donne la probabilité qu'un marcheur transite du sommet i vers le sommet j . Ainsi, on a $\mathbb{P}[X_k = j \mid X_{k-1} = i] = M(i, j) > 0 \iff [i, j] \in E(G)$.

1.3 Optimisation par Contraintes (CP)

En mathématiques discrètes est étudiée une branche particulière de la théorie de l'optimisation appelée *Optimisation combinatoire*. Cette discipline peut simplement être décrite comme

contenant l'ensemble des problèmes d'optimisation dont les variables de décision sont discrètes (choisies dans un ensemble dénombrable). Le terme combinatoire fait référence à la structure souvent combinatoire d'un problème de ce domaine, c'est-à-dire que le nombre de combinaison possible d'éléments de la structure peut être important. Un tel problème présente généralement un nombre de solutions admissibles qui augmente rapidement avec la taille du ou des arguments en entrée (tel que le nombre de sommets d'un graphe lorsque le graphe est un argument). Le terme *solution admissible* renvoie ici à une *solution*, une instantiation des variables de décision, qui respecte les contraintes. On parle aussi de l'espace des solutions ou de l'ensemble des solutions admissibles. Contrairement à plusieurs instances d'optimisation continue, il n'existe souvent pas de relations explicites entre les solutions admissibles d'un problème d'optimisation combinatoire comme la dérivée d'une fonction². Ainsi, la particularité des problèmes d'optimisation combinatoire est qu'il faut en pire cas énumérer chaque élément de l'espace des solutions et soit les parcourir un à un, soit retirer judicieusement des solutions jugées mauvaises tout en garantissant l'optimalité de la solution retenue.

La *Programmation par contraintes* (CP, de l'anglais « *Constraints Programming* ») est un paradigme de résolution de problèmes d'optimisation combinatoire³. On remarque que chaque problème d'optimisation se décrit comme l'optimisation d'une fonction objectif sous certaines contraintes. Ces contraintes encodent, avec la fonction objectif, l'information du problème d'optimisation et peuvent apparaître sous différentes formes. L'optimisation (programmation) par contraintes rend modulables les problèmes d'optimisation en nommant chaque contrainte séparément et en permettant de spécialiser chaque contrainte pour la rendre réutilisable dans d'autres problèmes. Un exemple classique est la contrainte ALLDIFFERENT(X_1, \dots, X_n) qui demande que les variables (X_1, \dots, X_n) soit toutes différentes. Cette contrainte n'est pas dépendante d'un problème particulier et peut être utilisée dans une grande variété de contextes différents, ce qui rend son étude plus intéressante.

Un programme par contraintes est composé d'un ensemble de variables, dites de décision, $\{X_1, \dots, X_n\}$. Chaque variable a son domaine de définition, soit l'ensemble des valeurs que la variable peut prendre. Plusieurs types de domaines existent, tels que les domaines énumérés dans lesquels chaque valeur possible est énumérée et les domaines bornés qui sont simplement des intervalles. On note que même dans le cas d'un domaine borné, seul un nombre dénombrable de valeurs peut être affecté à la variable. Sur ces variables sont finalement définies des contraintes, comme la contrainte ALLDIFFERENT. Une fonction objectif à optimiser est généralement ajoutée au problème, tel que mentionné plus haut. D'autres variables auxiliaires peuvent être définies pour simplifier la formulation du problème, mais la solution n'est décrite

2. Par exemple, pour minimiser la fonction $f(x) = x^2$ il convient simplement de dériver la fonction et de déterminer l'unique point où celle-ci s'annule. Au niveau algorithmique, la dérivée indique qu'il est toujours intéressant de continuer les calculs et dans quelle direction poursuivre. À l'inverse, les variables de décision d'un problème d'optimisation combinatoire sont généralement discrètes, la dérivée n'est donc pas définie.

3. Ce paradigme devrait aussi pouvoir être utilisé pour traiter des problèmes d'optimisation continue, mais est principalement utilisé pour traiter des problèmes discrets.

qu'avec les variables de décision.

Comme chaque variable a un domaine de définition, une tâche fondamentale de l'optimisation par contraintes est le *filtrage* des domaines, soit le retrait des valeurs des domaines qui contredisent une ou des contraintes. Ainsi, la contrainte ALLDIFFERENT a son propre algorithme de filtrage polynomial qui utilise la théorie des graphes. Cet algorithme s'exécute en quelques sortes indépendamment des autres algorithmes de filtrage des autres contraintes et fait partie de l'implémentation standard des logiciels de résolution de programmes par contraintes. Ainsi, cet algorithme peut être conçu, étudié puis optimisé indépendamment de tout problème spécifique d'optimisation combinatoire. Bref, l'intérêt de la programmation par contraintes est multiple : les programmes par contraintes peuvent s'écrire simplement comme une liste de contraintes nommées ; la programmation est modulable, chaque contrainte peut être perçue comme un module pouvant être amélioré séparément ; la modularité de ces contraintes permet la conception de catalogues de contraintes standards disponibles à la communauté et l'amélioration des performances de résolution d'un programme par contraintes peut être résumé par l'amélioration des performances d'un ou de quelques algorithmes de filtrages de certaines contraintes.

Chapitre 2

Théorie des jeux et jeux de poursuite

Ce chapitre introduit des concepts importants de la théorie des jeux et des jeux de policiers et voleurs. En premier lieu est présentée la manière dont les jeux de policiers et voleurs sont typiquement résolus, puis un lien entre cette méthode et une méthode analogue en théorie des jeux est illustré. Finalement, on présente les jeux stochastiques, qui seront réutilisés au chapitre 3.

2.1 Résolution de jeux finis

On dit qu'un jeu de policiers et voleurs est résolu lorsque la stratégie optimale des policiers est déterminée, en supposant que les voleurs jouent eux-mêmes optimalement. Ces jeux sont résolus à l'aide de deux méthodes distinctes, l'une algébrique et l'autre graphique. La première méthode est détaillée plus bas.

2.1.1 Résolution par relation d'ordre

Le jeu de policier et voleur classique est décrit dans l'article de Nowakowski et Winkler [39] ainsi que dans la thèse de doctorat de Quilliot [46]. Ceux-ci présentent un jeu se déroulant sur un graphe connexe fini (le cas infini est aussi étudié, on en trouve une description dans le livre de Bonato et Nowakowski [8]). Deux joueurs s'y opposent : un policier et un voleur. Ceux-ci, à tour de rôle, se déplacent de sommet en sommet le long des arêtes du graphe. Le jeu se déroule à temps discret. De plus, le policier est toujours le premier à jouer. Pour initialiser le jeu, le policier choisit un sommet de départ avant que le voleur ne choisisse le sien. Les joueurs se déplacent ensuite sur le graphe avec comme objectif pour le policier de capturer le voleur et pour ce dernier d'échapper à cette capture. Le policier remporte la partie lorsqu'il partage le même sommet que le voleur, peu importe le moment où ceci se produit.

La méthode présentée par Nowakowski et Winkler pour résoudre ce jeu, c'est-à-dire pour déterminer la stratégie optimale du policier, définit une relation de préordre pour comparer

des sommets du graphe. Quoique très astucieuse, cette méthode peut être perçue comme une réinterprétation d'une méthode classique développée quelque trente années auparavant pour résoudre, sous certaines conditions, des problèmes généraux de la théorie des jeux. Ainsi, pour faire le parallèle entre les jeux de policiers et voleurs et la théorie des jeux, on présente ici la méthode de Nowakowski et Winkler ainsi que son homologue tirée de la théorie des jeux.

Les prochains résultats ne sont pas démontrés ici, mais on renvoie au livre de Bonato et Nowakowski [8] pour leurs démonstrations et plus d'explications.

On définit premièrement la relation d'ordre \preceq_n , objet central de l'article de Nowakowski et Winkler. Celle-ci est définie de telle sorte que deux sommets, x et y , sont en relation $x \preceq_n y$ si et seulement si le policier positionné en y peut capturer en n tours le voleur se tenant en x .

Définition 2.1.1. *Soit G le graphe de jeu et $x, y \in V(G)$ deux sommets. On pose*

$$\begin{aligned} x \preceq_0 y &\iff x = y, \\ x \preceq_n y &\iff \forall v \in N[x], \exists u \in N[y] : v \preceq_{n-1} u. \end{aligned} \tag{2.1}$$

Dans l'expression (2.1), n représente le nombre de tours nécessaires à la capture du voleur. Cette relation est définie pour déterminer si une position y permet au policier de gagner contre un voleur situé en x en n tours. Ainsi, le seul moyen qu'un policier en y gagne contre un voleur sur x en 0 tours est, par définition du jeu, que les deux joueurs soient sur le même sommet, c'est-à-dire que $x = y$. Autrement, $n > 0$ et supposons que c'est au voleur à jouer. Alors, le policier doit chercher une position $u \in N[y]$ lui garantissant que peu importe l'action $v \in N[x]$ du voleur, le policier gagne en $n - 1$ tours. L'ingéniosité de l'expression (2.1) est donc sa structure récursive : pour gagner en $n - 1$ tours, le policier doit vérifier si $v \preceq_{n-1} u$, ce qui est déjà défini.

Exemple 2.1.2. Illustrons la relation \preceq_n sur les trois graphes de la figure 2.1. Le policier est toujours assuré de gagner sur ceux-ci. Sur le graphe 2.1(a), on remarque par inspection que les sommets A, B, C ont tous pour propriété que $x \preceq_1 y$, avec $x, y \in \{A, B, C\}$. En effet, pour deux sommets distincts $x, y \in \{A, B, C\}$ on a clairement $x \not\preceq_0 y$. Cependant si un voleur se tenait sur $x \in \{A, B, C\}$ et un policier sur $y \in \{A, B, C\} \setminus \{x\}$, alors peu importe le prochain sommet choisi par le voleur il se retrouverait nécessairement sur la même position que le policier, sauf s'il restait sur place. Autrement dit, pour tous sommets x, y , on a forcément $\forall v \in N[x], \exists u \in N[y] : v = u$.

Le second graphe, le graphe 2.1(b)¹, n'a pas la même propriété que le premier. Comme celui-ci contient encore le triangle 2.1(a), les trois sommets A, B, C sont en relation $x \preceq_1 y$ pour tout $x \in \{A, B, C\}$ et $y \in \{A, B, C\}$. Par contre, on voit que $C \preceq_2 D$. En effet, lorsque le policier se déplace en C , alors peu importe l'action $x \in \{A, B, C\}$ du voleur on a forcément $x \preceq_1 C$ puisque les deux joueurs se trouvent se le triangle $\{A, B, C\}$. En suivant le même

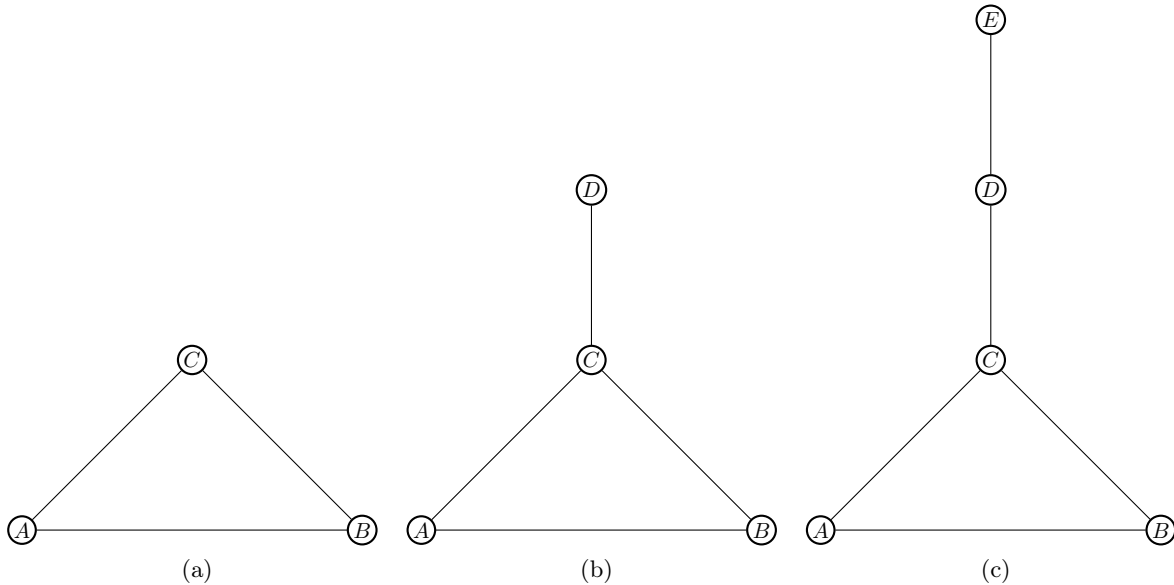


FIGURE 2.1 – Trois graphes policiers-gagnants.

raisonnement, on voit que $D \preceq_3 E$ sur le graphe 2.1(c). Autrement dit, sur le second graphe le policier doit assurément faire deux mouvements pour gagner et trois sur le dernier.

Exemple 2.1.3. On illustre la construction de la relation (2.1) sur un jeu plus connu, le jeu des échecs. Bien que celui-ci ne soit pas un jeu de policier et voleur, les idées derrière la relation peuvent être reprises.

Ainsi, supposons qu'on veuille vérifier si les blancs peuvent gagner aux échecs contre un opposant, les noirs, jouant de manière optimale. Comme le jeu a un nombre fini de cases, il existe un nombre fini, quoique extrêmement élevé, de manières pour les blancs de gagner. Les blancs peuvent faire une liste de chaque configuration les faisant gagner et définir la relation $x \preceq_0 y$ si x est la position des noirs et y celle des blancs dans cette configuration. Ensuite, les blancs doivent vérifier s'il est possible d'en arriver à une telle configuration lorsque les noirs jouent de manière optimale. Pour y arriver, il se doit qu'à l'avant-dernier tour les blancs soient en mesure de forcer les noirs à entrer dans une configuration $x \preceq_0 y$. Autrement dit, lorsque les noirs sont sur le point de jouer, peu importe l'action x qu'ils envisagent, les blancs doivent pouvoir jouer une action y les favorisant au dernier tour. Bref, on a construit la deuxième partie de la relation, $x \preceq_1 y$ si et seulement si pour chaque action v il existe une réponse u telle que $v \preceq_0 u$. En appliquant successivement ce raisonnement, on obtient la relation générale $x \preceq_n y$.

Un résultat qui est généralement mentionné sans plus d'explications, mais qui est d'une importance considérable est le suivant. On note préalablement que, comme les joueurs peuvent

1. Les graphes 2.1(b) et 2.1(c) sont en fait des exemples de graphes dits suçons (en anglais *lollipop*). Ceux-ci sont construits en adjoignant un chemin P_n et un graphe complet K_m à une de ses extrémités.

chacun rester sur place, la relation $x \preceq_n y$ est incluse dans la relation $x \preceq_{n+1} y$. Autrement dit, $\{x, y \in V(G) \mid x \preceq_n y\} \subseteq \{x, y \in V(G) \mid x \preceq_{n+1} y\}$.

Lemme 2.1.4. *Il existe un nombre $k \in \mathbb{N}$ tel que, pour toute paire de sommets $x, y \in V(G)$, $x \preceq_k y = x \preceq_{k+1} y$. Autrement dit, la relation devient stationnaire.*

Ce résultat est fondamental puisqu'il justifie le calcul de cette relation. En effet, comme celle-ci se calcule récursivement, cette stationnarité garantit qu'en temps fini il est possible de comparer deux sommets d'un graphe. Qui plus est, lorsque le graphe est fini il est assez facile d'en conclure que toute paire de sommets peut maintenant être comparée en temps fini. Ainsi, puisque la relation (2.1) devient stationnaire à partir d'un indice n_0 , on écrit simplement $x \preceq y$ pour signifier $x \preceq_{n_0} y$.

Le résultat le plus important au niveau des jeux de policiers et voleurs, justifiant même la définition de la relation (2.1) est le suivant. On parle de *stratégie*, de manière assez générale, comme d'un plan permettant en toute occasion à un joueur de déterminer sa prochaine action à jouer. Une *stratégie gagnante*, contre un ou plusieurs joueurs, est une stratégie garantissant au joueur qui l'emploie de gagner contre ses adversaires, peu importe la ou les stratégies jouées par ces derniers. Pour plus de détails, on renvoie encore à la section 2.2 sur les jeux stochastiques.

Théorème 2.1.5. *Il existe une stratégie gagnante pour le policier si et seulement s'il existe un sommet $y^* \in V(G)$ tel que pour tout $x \in V(G)$, $x \preceq y^*$.*

Remarque 2.1.6. *On peut remarquer deux aspects importants du théorème précédent.*

1. *S'il existe une stratégie gagnante pour le policier, alors (par définition) peu importe la stratégie employée par le voleur celui-ci est assuré de perdre. Inversement, si une telle stratégie n'existe pas, le voleur peut gagner toute partie jouée dans ce jeu. Le voleur a donc une stratégie gagnante dans ce dernier cas.*
2. *La stratégie gagnante est entièrement dépendante du graphe sur lequel le jeu est joué. En effet, l'existence d'une stratégie gagnante sur un graphe ne garantit rien quant à l'existence d'une stratégie gagnante sur un autre graphe. En termes informatiques, le graphe fait partie de l'entrée. Cette remarque mène les chercheurs à parler de graphes policiers-gagnants.*

2.1.2 Le point de vue de la théorie des jeux, le théorème de Kuhn

Dans les années 1930, Zermelo (aussi immortalisé pour avoir contribué aux axiomes Zermelo-Frankel-axiome du choix) démontra que, sous certaines hypothèses, dans le jeu des échecs, soit il existe assurément une stratégie gagnante pour les blancs, soit cela est vrai pour les noirs, soit les joueurs peuvent toujours faire partie nulle [55]. Ce résultat allait quelques années plus tard être la pierre angulaire de ce qu'on nomme aujourd'hui les *jeux combinatoires*, dont font

partie les jeux de policiers et voleurs [6]. Cependant, la théorie des jeux en général aussi allait profiter de ce résultat et Harold Kuhn, vers 1950 [41, 53], démontra une généralisation de celui-ci dans le cadre général de la théorie des jeux. La démonstration de Kuhn repose sur une méthode aujourd'hui connue sous le nom *d'induction à rebours* et qui est d'une importance capitale dans plusieurs domaines. Le théorème de Kuhn est présenté avec une esquisse de sa démonstration.

La formulation typique de ce théorème est quelque peu altérée pour simplifier l'exposé et pour garder un vocabulaire accessible. On note qu'un jeu fini est un jeu dans lequel chaque joueur n'a accès qu'à un nombre fini de positions et ne fait qu'un nombre fini de mouvements par tour (comme le *Tic-Tac-Toe*). Un jeu alterné se joue à tour de rôle et un jeu à information parfaite, encore comme le *Tic-Tac-Toe*, a la propriété que chaque joueur possède toute l'information nécessaire à sa prise de décision, incluant les règles complètes du jeu ainsi que l'information sur les coups précédemment joués par les autres joueurs.

Théorème 2.1.7. *Dans tout jeu fini, alterné et à information parfaite, soit un joueur possède une stratégie gagnante, soit tous les joueurs ont le moyen de faire une partie nulle.*

Schéma de la démonstration. L'idée principale de cette démonstration, l'induction à rebours, est illustrée dans l'exemple 2.1.3 qui précède. Comme le jeu est fini, on peut écrire T pour le nombre maximal de tours joués dans une partie. Supposons que dans un jeu fini, alterné et à information parfaite on veuille vérifier qu'un joueur peut gagner une partie peu importe les mouvements des autres joueurs. Tel que dans l'exemple 2.1.3, comme le jeu est fini il est possible pour le joueur, nommons-le Bob, de faire une liste des configurations de jeu dans lesquelles il gagne la partie puisqu'il y en a un nombre limité. Nommons l'ensemble de ces configurations F . Pour simplifier, supposons que Bob n'ait qu'un seul adversaire et nommons-la Alice. Au tour $T - 1$, si Bob a une stratégie gagnante alors peu importe l'action de son adversaire il est certain de gagner. Autrement, soit Alice a une stratégie gagnante empêchant Bob de la vaincre, soit ce n'est le cas de ni un ni l'autre et les deux font partie nulle. Par induction, on déduit que cette propriété est aussi vraie aux tours $T - 2, T - 3, \dots, 0$. Ainsi, soit un des joueurs possède une stratégie gagnante en début de partie lui assurant la victoire, soit les deux joueurs font partie nulle. \square

En ce qui nous concerne, l'importance du théorème 2.1.7 vient de sa démonstration, que l'on peut retrouver sous une autre forme dans le livre de Osborne et Rubinstein [41].

2.2 Les jeux stochastiques

La théorie des jeux stochastiques vient consolider une partie de la théorie des jeux traditionnels en permettant de modéliser l'ajout d'événements aléatoires dans des jeux. Cette branche

de la théorie des jeux est attribuable à Shapley [50] qui la décrivait dans les années 1950. Shapley décrivait des jeux simultanés (les joueurs jouent en même temps) et répétés (les joueurs s'affrontent dans une suite de parties). Dans ces jeux, chaque partie se compose d'une action simultanée de chaque joueur, après quoi les joueurs reçoivent un certain gain, valeur qu'ils doivent maximiser, puis le jeu transite aléatoirement vers une autre partie.

Exemple 2.2.1. On peut imaginer le jeu suivant, qui est une variante simple d'un jeu de poker. À chaque tour, chaque joueur se voit distribuer cinq cartes qui sont visibles à tous. Ils peuvent demander de remplacer un certain nombre de ces cartes une seule fois. Ensuite les joueurs doivent agencer celles-ci pour tenter de former des mains de poker (paire, triple, suite, etc.). Le joueur avec la plus haute main reçoit un gain proportionnel à la valeur de sa main. Un joueur peut décider de garder les cartes qui ne lui ont pas servi pour le prochain tour, auquel cas il se verra attribuer moins de cartes au prochain tour. Chaque joueur doit en tout temps ne posséder que cinq cartes. On peut voir la distribution des cartes comme une transition aléatoire d'un tour vers un autre tour. Cette transition dépend des actions que les joueurs ont exécutées, un joueur décidant de garder deux cartes se verrait attribuer trois cartes aléatoirement au tour suivant. Les transitions aléatoires sont donc conditionnelles aux actions des joueurs. Les stratégies des joueurs dépendent donc de la qualité de la main qu'ils possèdent au tour présent ainsi que la probabilité d'obtenir une bonne main aux tours subséquents.

Une définition moderne, et plus facilement applicable au contexte des jeux sur des graphes, de la théorie des jeux stochastiques est présentée par Michael Ummels [56] comme le fruit de ses recherches doctorales à l'université d'Aachen en Allemagne. C'est cette définition qui est présentée dans ce mémoire.

Un jeu stochastique est un jeu alterné. Le jeu alterne entre l'action d'un joueur, une action aléatoire qui n'est contrôlée par personne et l'action d'un autre joueur, distinct du premier. Les joueurs, en nombre fini, sont choisis dans l'ensemble $\Pi = \{0, \dots, n-1\}$. Ils s'affrontent sur un graphe dirigé (faiblement²) connexe $D = (V, E)$. Chacun d'entre eux a le contrôle sur un sous-ensemble $V_i \subset V$, avec $i \in \Pi$, des sommets de D duquel il peut choisir le prochain sommet à visiter, ce qui constitue son action. Les actions jouées forment un chemin infini qui constitue une *partie*. Il est possible que le graphe D soit fini, mais on accepte qu'un chemin repasse plus d'une fois sur un même sommet, ce qui permet l'existence de chemins infinis.

La stochasticité du jeu provient des sommets aléatoires $V_s \subseteq V$. En effet, après chaque action d'un joueur le jeu transite vers un sommet aléatoire d'où le prochain sommet est choisi au hasard selon une certaine distribution. On peut imaginer que les sommets aléatoires de V_s sont contrôlés par la *nature* qui emploie une stratégie inconnue des joueurs.

2. En théorie des graphes, un graphe dirigé est faiblement connexe s'il est connexe lorsque les directions des arêtes sont retirées. De même, un graphe est fortement connexe s'il est connexe aussi en tenant compte des directions.

La définition suivante 2.2.2 formalise le concept de jeux stochastiques.

Définition 2.2.2. *Un jeu stochastique est caractérisé par un tuple*

$$\mathcal{G} := (\Pi, V, (V_i)_{i \in \Pi}, \Delta, \chi, (W_i)_{i \in \Pi}).$$

V est l'ensemble des sommets d'un graphe dirigé (faiblement) connexe D . $\Pi = \{0, \dots, n-1\}$ est l'ensemble fini des joueurs de \mathcal{G} . Chaque joueur $i \in \Pi$ contrôle un ensemble de sommets V_i alors qu'un ensemble complémentaire $V_s = V \setminus \bigcup_{i \in \Pi} V_i$ contient les sommets stochastiques. Les ensembles V_i sont disjoints pour chaque $i \in \Pi$. Un jeu est dit fini si V est fini.

Une relation de transition $\Delta \subseteq V \times ([0, 1] \cup \{\perp\}) \times V$ décrit les transitions, contrôlées et stochastiques, le long des arêtes de D . Le symbole \perp dénote une transition débutant d'un sommet contrôlé. Si v est un sommet stochastique et w un sommet quelconque, alors il existe $p \in [0, 1]$ qui est l'unique nombre tel que $(v, p, w) \in \Delta$. On nomme p la probabilité de transition de v à w (on a donc $\sum_{w \in V: (v, p, w) \in \Delta} \Delta(v, p, w) = 1$). On a toujours, pour $u, v \in V$, que $(u, \perp, v) \in \Delta$ ou $(u, p, v) \in \Delta$ seulement si $v \in N[u]$.

Une fonction de coloriage $\chi : V \rightarrow C$ permet de caractériser les objectifs des joueurs. Soit C un ensemble de couleurs et C^ω l'ensemble des suites infinies de C . L'ensemble $W_i \subseteq C^\omega$ du joueur $i \in \Pi$ encode l'objectif du joueur i . Chaque joueur doit donc parcourir une suite infinie de couleurs.

Il est toujours possible de décrire un jeu à l'aide d'une description exhaustive de ses configurations possibles. Par exemple, le jeu d'échecs n'admet qu'une configuration possible au tour initial, avant que le jeu ne commence, mais 400 configurations possibles au premier tour (soit 20 mouvements possibles pour chaque joueur). Chaque configuration est unique et décrit exactement les positions des pièces de chaque joueur sur le plateau de jeu. Dans le cas d'un jeu stochastique, ces configurations sont contenues dans les ensembles de sommets contrôlés $(V_i)_{i \in \Pi}$. Ainsi, lorsqu'un joueur décide de quitter un sommet contrôlé pour se diriger vers un sommet voisin, il change la configuration de jeu courante pour tenter d'obtenir une configuration l'avantageant. Une partie n'est donc qu'une suite de changements de configurations de jeu admissibles.

La relation Δ permet de définir les successeurs d'un sommet, l'ensemble des sommets accessibles à partir de n'importe quel sommet. Elle décrit donc implicitement l'ensemble des arêtes du graphe D .

Le coloriage du graphe D sert à décrire les objectifs que doit atteindre chaque joueur pour gagner une partie. Aux échecs, par exemple, toute configuration de jeu dans laquelle l'adversaire est échec et mat permet au joueur de remporter la partie. Pour reconnaître ces configurations, on les identifie à l'aide de couleurs. On colorie les configurations admissibles où aucun

joueur ne gagne en rouge. Les configurations dans lesquelles les blancs gagnent, c'est-à-dire dans lesquelles les noirs sont échec et mat, sont coloriées en blanc et inversement pour les configurations gagnantes pour les noirs qui sont coloriées en noir. Les configurations restantes sont coloriées en gris. L'objectif des blancs serait donc simplement de trouver un chemin sur le graphe D leur assurant de traverser au moins un sommet blanc.

Comme C^ω est infini, il est possible que W_i (l'ensemble des chemins permettant au joueur i de gagner) le soit aussi. Autrement dit, il est possible qu'un joueur doive répéter certaines configurations de jeu un nombre infini de fois pour gagner une partie. Par exemple, un voleur gagne si et seulement si il lui est possible de toujours échapper au policier. Cependant, un seul type d'objectif nous intéresse dans ce mémoire et on ne décrit en détail que celui-ci : l'accessibilité simple. Lorsque aucun joueur ne peut atteindre son objectif, il est nécessaire de déterminer quand une partie est considérée comme nulle, mais cela n'est pas étudié dans ce mémoire puisqu'il est impossible pour les joueurs de réaliser une partie nulle dans un jeu de policiers et voleurs.

L'objectif des joueurs dans un jeu d'accessibilité est d'atteindre certaines configurations de jeu particulières. Dans le cas où une configuration est atteinte, le joueur dont il était l'objectif de l'atteindre est déclaré gagnant. Cet objectif simple permet de représenter un éventail très large de jeux, dont les échecs et les jeux de policiers et voleurs. En effet, la capture d'un voleur se représente par la transition vers une configuration de jeu dans laquelle les voleurs sont capturés. Dans cette représentation il est clair que le policier doit atteindre cet état de jeu, il a donc un objectif d'accessibilité³.

Les jeux stochastiques sont considérés comme des généralisations des processus de Markov décisionnels (de l'anglais « *Markov Decision Processes* », MDP) [3, 45]. Eux-mêmes généralisent la chaîne de Markov.

Exemple 2.2.3 (La chaîne de Markov). Du point de vue des jeux stochastiques, une chaîne de Markov n'est qu'une suite infinie de sommets stochastiques. Aucun joueur ne contrôle de sommets du graphe dirigé D . Soit S l'ensemble d'états de la chaîne. Alors on pose $V = V_s = S$. De la relation Δ , on déduit que le triplet $(v, p, w) \in \Delta$ donne la probabilité de transition $\mathbb{P}[w | v] = p$. Comme aucun joueur n'agit sur le graphe, il n'est pas nécessaire de définir d'objectif.

Le processus de Markov décisionnel peut aussi être interprété comme un cas spécifique d'un jeu

3. Il existe d'autres objectifs plus complexes que les objectifs d'accessibilité tels que ceux dits de Büchi. Ceux-ci requièrent du joueur correspondant que le chemin qu'il parcourt lui fasse traverser infiniment souvent un ensemble de couleurs lui étant associées. Dans le cas des jeux de policiers et voleurs, on pourrait représenter l'objectif des voleurs par un objectif de Büchi : chaque configuration dans laquelle les voleurs sont en vie est coloriée, disons, en rouge et l'objectif des voleurs est simplement de parcourir un chemin contenant un nombre infini de sommets rouges. Dans ce cas, l'objectif des voleurs est complémentaire à l'objectif du policier. Si le policier gagne en traversant un sommet bleu, alors on aurait $W_v = C^\omega \setminus W_p$, où v et p dénotent les voleurs et les policiers.

stochastique à un seul joueur. Cependant, la correspondance n'est pas présentée ici puisqu'elle nécessite une redéfinition des objectifs de l'agent dans un MDP pour cadrer avec le modèle de jeux stochastiques présenté par Ummels [56]. Cette correspondance dépasse le cadre de ce mémoire. Malgré tout, d'un point de vue conceptuel le MDP décrit un processus markovien dans lequel un agent prend des décisions pour contrôler les transitions, ce qui correspond au jeu stochastique à un seul joueur.

Dans un jeu stochastique, comme dans un jeu général, les joueurs font appel à des stratégies. Soit \mathcal{P} l'ensemble des chemins de D et \mathcal{P}^{V_i} l'ensemble des chemins de D se terminant en un sommet de V_i , peu importe le sommet de départ. Une stratégie pure, ou déterministe, pour le joueur i est une fonction $\sigma : \mathcal{P}^{V_i} \rightarrow V$ associant à chaque chemin terminant en un sommet contrôlé par i un sommet de V . Un profil de stratégies $\bar{\sigma} = (\sigma_i)_{i \in \Pi}$ regroupe les stratégies des joueurs. Comme on parle généralement de jeux de policiers et voleurs, donc des jeux à deux joueurs, un profil de stratégie s'écrit ici comme un couple (σ, τ) .

Les stratégies utilisées plus loin, aux chapitres 3 et 4, ne dépendent que du sommet courant (et non du chemin parcouru précédemment). Dans le jargon des jeux stochastiques, on dit que ces stratégies sont *stationnaires*. Au chapitre 3 on utilisera aussi le terme stationnaire pour désigner un autre concept, mais le contexte va permettre de différencier entre ces deux notions.

On peut aussi définir un sous-jeu d'un jeu général \mathcal{G} d'une manière assez naturelle. Soit $x \in \mathcal{P}$ un chemin de D correspondant à un début de partie. Alors $\mathcal{G}[x]$ est un jeu résiduel débutant au dernier sommet de x , autrement dit $\mathcal{G}[x]$ est le jeu restant lorsque x a déjà été parcourue. Formellement, un jeu résiduel ne diffère d'un jeu normal que sur ses objectifs : on précise dans un jeu résiduel $\mathcal{G}[x]$ que les objectifs des joueurs sont fonctions de la suite de sommets parcourus x , autrement dit l'objectif du joueur i devient $\{\alpha \in C^\omega : \chi(x)\alpha \in W_i\}$, $\forall i \in \Pi$. Certains objectifs sont dits indépendants du préfixe lorsque l'ajout d'un préfixe, c'est-à-dire la concaténation à gauche d'une suite finie de sommets, à un jeu n'altère pas les conditions de victoire des joueurs. Dans ces jeux les joueurs peuvent gagner à n'importe quel moment du jeu. Dans les jeux indépendants du préfixe, il est inutile de parler de jeu résiduel puisque pour tout préfixe fini x , on a $\mathcal{G}[x] = \mathcal{G}$, c'est-à-dire un jeu indépendant du préfixe correspond à n'importe quel de ses jeux résiduels.

2.2.1 Valeurs et stratégies optimales

On dit d'une stratégie qu'elle est optimale si elle maximise ce qu'on nomme la *valeur d'un jeu*. Cette valeur pour le joueur i , notée $\text{val}_i^{\mathcal{G}}$, est la probabilité maximale que ce joueur gagne une partie \mathcal{G} en s'opposant à un ou plusieurs adversaire(s) jouant eux-mêmes de manière optimale. Pour simplifier, on ne définit que cette notion pour les jeux à deux joueurs puisque seuls ceux-ci nous intéressent.

Comme chaque joueur i doit prendre en considération les stratégies que ses adversaires peuvent

employer, on commence par définir la probabilité minimale de victoire de i contre n'importe quelle stratégie valide. Ainsi, étant donné une stratégie σ pour le joueur i et un sommet de départ v , on écrit

$$\text{val}_i^\sigma(v) := \inf_{\tau} \mathbb{P}[W_i \mid \tau, \sigma, v]$$

pour signifier la probabilité minimale de gagner pour le joueur i lorsque son adversaire peut utiliser n'importe quelle stratégie τ ⁴. La probabilité $\mathbb{P}[W_i \mid \tau, \sigma, v]$ est perçue ici comme la probabilité que le joueur i atteigne son objectif W_i en usant de la stratégie σ , alors que son adversaire utilise la stratégie τ , le tout en commençant à un sommet v .

La valeur d'un jeu pour le joueur i est simplement la probabilité qu'il gagne en jouant sa meilleure stratégie possible contre un adversaire intelligent,

$$\text{val}_i^{\mathcal{G}}(v) := \sup_{\sigma} \inf_{\tau} \mathbb{P}[W_i \mid \tau, \sigma, v]$$

Cette valeur est définie à partir de n'importe quel sommet v . En particulier, elle est définie pour le sommet de départ du jeu v_0 .

Remarque 2.2.4. *Dans la théorie des jeux, en général la valeur d'un jeu décrit plutôt le gain maximal qu'un joueur peut s'attendre à obtenir en s'opposant à des adversaires qui jouent de manière optimale. Ici, cette valeur est une probabilité entre autres puisque aucun gain à proprement parler n'est défini : un joueur gagne s'il atteint son objectif, mais aucun joueur ne reçoit de valeur numérique qu'il cherche à optimiser. D'un autre côté, de par la nature stochastique du jeu \mathcal{G} , il est naturel de se demander quelle est la probabilité qu'un joueur gagne et, ce faisant, de définir ainsi la valeur d'un jeu.*

Il est à noter qu'on utilise ici un infimum et un supremum, au lieu d'un minimum et d'un maximum. C'est un détail technique qu'il est important de considérer. Le minimum (maximum) d'un ensemble fait partie de cet ensemble, alors que l'infimum (supremum) peut être à l'extérieur. Formellement, l'infimum (supremum) est simplement la plus grande (petite) valeur bornant inférieurement (supérieurement) un ensemble. En théorie des jeux, on doit montrer que la valeur d'un jeu existe après l'avoir définie. Dans certains jeux, cette valeur n'existe pas, autrement dit l'infimum (supremum) n'est pas atteint et on ne peut pas utiliser la notion de minimum (maximum).

Maintenant que la valeur d'un jeu est définie, on peut finalement parler de stratégie optimale. Une stratégie optimale pour le joueur i est simplement une stratégie σ atteignant la valeur du jeu pour ce joueur. Autrement dit, la stratégie σ est dite *optimale* si et seulement si

$$\text{val}_i^{\mathcal{G}} = \text{val}_i^\sigma.$$

4. Dans les jeux à plus de deux joueurs, on peut remplacer τ , sous l'infimum, par un profil de stratégies σ_{-i} contenant l'ensemble des stratégies des joueurs autres que i .

La stratégie optimale est donc exactement la stratégie garantissant la plus grande probabilité de victoire à un joueur s'opposant à un adversaire jouant aussi optimalement.

Pour qu'une stratégie soit optimale, il est donc nécessaire que la valeur du jeu existe. On dit donc d'un jeu qu'il est *déterminé* si sa valeur existe. Le fait qu'un jeu soit déterminé n'est qu'une condition nécessaire à l'existence de stratégies optimales, mais n'est pas suffisante. En effet, pour qu'un jeu soit déterminé il doit y exister une valeur v telle que $v = \sup_{\sigma} \inf_{\tau} \mathbb{P}[W_i \mid \tau, \sigma, v]$ pour chaque joueur i . Ceci n'implique pas directement que les stratégies optimales existent puisque celles-ci peuvent être à l'extérieur des ensembles de stratégies des joueurs. On note que dans un jeu à deux joueurs la valeur d'un joueur est le complémentaire de la valeur de son adversaire,

$$\text{val}_0^{\mathcal{G}} = 1 - \text{val}_1^{\mathcal{G}}.$$

La valeur d'un jeu peut être difficile à calculer, incidemment la stratégie optimale peut être difficile à déterminer. Cependant, la valeur d'un jeu se prête aisément à l'approximation. On dit d'une stratégie σ qu'elle est ϵ -optimale si et seulement si

$$\text{val}_i^{\mathcal{G}} \geq \text{val}_i^{\sigma} - \epsilon.$$

Autrement dit, une stratégie ϵ -optimale atteint une valeur qui est à ϵ près de la valeur du jeu. Ce concept est important dans plusieurs applications, mais n'est que mentionné avec quelques résultats dans ce mémoire.

Maintenant que les définitions importantes ont été faites, on peut mentionner certains résultats importants de la théorie des jeux stochastiques. Ceux-ci sont présentés sans démonstration, mais plus d'informations peuvent être trouvées dans les travaux de Ummels [56].

Théorème 2.2.5 (Martin ; Maitra et Sudderth [32, 35]). *Tout jeu stochastique à deux joueurs avec objectifs d'accessibilité⁵ est déterminé. De plus, pour tout $\epsilon > 0$, chaque joueur possède une stratégie ϵ -optimale pure.*

Ainsi, selon ce théorème tous les jeux de policiers et voleurs sont déterminés. On montre en effet au chapitre 3 et à l'annexe A que les jeux de policiers et voleurs peuvent s'écrire comme des jeux stochastiques. L'existence de stratégies ϵ -optimales nous garantit qu'il est toujours possible d'obtenir des stratégies qui sont au moins à ϵ près de la valeur optimale. On verra plus loin que c'est une garantie un peu faible, cependant, puisque l'existence de stratégies optimales est aussi garantie dans nos jeux.

5. Martin, Maitra et Sudderth présente en fait un résultat beaucoup plus fort en démontrant celui-ci pour tout objectif dit *borélien*. Cependant, le concept d'ensemble de Borel, ou borélien, provient de la théorie de la mesure et cette théorie dépasse grandement le cadre de ce mémoire. On ne présente donc pas de théorie de la mesure ici. Ceci se fait cependant sans perte de généralité pour les résultats qui suivent puisque tous les jeux considérés plus loin ont des objectifs d'accessibilité. On garantit donc au lecteur qu'un objectif d'accessibilité forme un ensemble de Borel.

On peut aussi présenter un résultat plus fort dans le contexte de certains jeux de poursuite qui sont indépendants de préfixe, c'est-à-dire dans lesquels les policiers peuvent gagner à tout moment. Ces jeux nécessitent entre autres qu'il n'y ait aucune limite de temps et n'incluent donc pas le jeu de l'OSP (présenté en détail au chapitre 4).

Théorème 2.2.6 (Gimbert et Horn [19]). *Dans tout jeu stochastique fini à deux joueurs avec objectifs indépendants de préfixe il existe des stratégies résiduelles optimales pures pour les deux joueurs.*

Dans le théorème de Gimbert et Horn, comme les jeux sont supposés indépendants de préfixe (et que $\mathcal{G} = \mathcal{G}[x]$) il existe non seulement des stratégies optimales résiduelles, mais aussi des stratégies optimales globales.

Finalement, on dit souvent que les joueurs d'un jeu de policiers et voleurs jouent optimalement en signifiant qu'ils jouent *déterministiquement* optimalement. Autrement dit, il est souvent supposé que la stratégie optimale est pure (déterministe), voire stationnaire. Il est cependant possible de démontrer ce résultat formellement.

Théorème 2.2.7 (Liggett et Lippman [31]). *Pour tout jeu stochastique fini à deux joueurs avec objectifs d'accessibilité, chaque joueur possède une stratégie optimale déterministe et stationnaire.*

Dans le texte de Ummels [56], ce résultat est présenté sous la forme d'un corollaire et les objectifs du jeu ne sont pas d'accessibilité. Cependant, le corollaire présenté est plus fort que le théorème précédent, qui est suffisant dans le cadre de ce mémoire. Ce dernier théorème est donc le plus important des trois puisqu'il garantit non seulement que les joueurs peuvent jouer optimalement, mais qu'ils cherchent à jouer selon des fonctions qui ne dépendent que du sommet courant.

Finalement, on ne présente que quelques résultats de complexité algorithmique, puisque la plupart des résultats présentés par Ummels [56] dépassent le cadre de ce mémoire. De plus, ces résultats sont en quelques sortes ambigus. En effet, on retrouve beaucoup de résultats formulés avec les classes NP, coNP et NP-complet, ce qui est quelque peu douteux lorsque l'on sait que certaines instances de jeux de policiers et voleurs peuvent être de taille exponentielle selon la taille du graphe (comme les jeux k policiers et voleurs). En effet, pour résoudre les jeux à k policiers, on définit un nouveau graphe avec $|V(G)|^k$ sommets sur lequel les policiers se déplacent. Dans ce cas, on peut définir le graphe dirigé D avec le graphe $|V(G)|^k$ et calculer une complexité algorithmique polynomiale selon la taille de D en outrepassant le fait que cette complexité soit exponentielle (pour k non fixé) selon la taille de G . On doit donc faire attention avec ces résultats et garder à l'esprit que ceux-ci sont formulés selon la taille du graphe D et non pas selon la taille du graphe de départ sur lequel se déroule le jeu, par exemple, de policiers et voleurs.

Théorème 2.2.8. *Le problème dit quantitatif est le suivant.*

Soit \mathcal{G} un jeu stochastique à deux joueurs, v un sommet et $p \in [0, 1] \cap \mathbb{Q}$ un nombre rationnel. Décider si $\text{val}_0^{\mathcal{G}}(v) \geq p$.

Ce problème est dans $\text{NP} \cap \text{coNP}$.

On réitère qu'il faut être prudent par rapport à l'appartenance à $\text{NP} \cap \text{coNP}$ du problème précédent. La complexité du problème quantitatif est évaluée selon la taille du jeu \mathcal{G} qui est, elle, caractérisée par la taille du graphe D . Malgré tout, cette question va revenir plus tard dans le chapitre 3 où sera présentée une méthode pour résoudre une grande classe de jeux de policiers et voleurs. La complexité de cette méthode est évaluée dans ce chapitre et on verra qu'il est possible que certains jeux ne soient pas résolubles en temps polynomial. Le théorème précédent vient appuyer cette intuition, sans pour autant la confirmer.

Chapitre 3

Généralisation des jeux de poursuite stochastiques

3.1 Introduction

Les jeux de policiers et voleurs sont étudiés comme exemples de jeux de poursuite à temps discret sur des graphes depuis la publication de la thèse de doctorat de Quiliot [46] en 1978 et, indépendamment, de l'article de Nowakowski et Winkler [39] en 1983. Les deux publications décrivent un jeu se déroulant à tour de rôle et dans lequel un seul policier poursuit un voleur sur les sommets d'un graphe. Le tout se déroule à temps discret et avec information parfaite. Le policier gagne s'il en vient éventuellement à partager le même sommet que le voleur, autrement ce dernier remporte la partie. Ce jeu est complètement déterministe puisqu'on y suppose que les joueurs jouent de manière optimale. Ainsi, puisque le jeu se déroule à information parfaite, il est toujours vrai qu'au moins un des joueurs a une stratégie gagnante déterministe. Ce point, qui a déjà été présenté au chapitre 2 avec quelques définitions et résultats de la théorie des jeux, sera réexaminé plus tard.

Depuis la première description du jeu de policiers et voleurs, plusieurs variantes ont fait surface. Aigner et Fromme [1] ont notamment présenté en 1984 le *nombre de policiers* (que l'on nomme par la suite par son appellation anglaise *cop number*) comme un paramètre de graphes décrivant le nombre minimal de policiers requis sur un graphe pour capturer le voleur. À partir de là, plusieurs variantes ont été décrites, chacune modifiant un ou plusieurs paramètres du jeu tels que la vitesse des joueurs, le nombre de policiers, le rayon de capture des policiers, etc. On renvoie au livre de Bonato et Nowakowski [8] pour une description de ces différentes variantes. La revue de littérature de Fomin et Thilikos [17] est aussi une bonne source traitant des problèmes de fouilles de graphe garanties, dont les jeux de policiers et voleurs font partie. Dans les problèmes de fouille de graphes, l'objectif est toujours de retrouver un objet perdu sur un graphe. Les problèmes dits de fouille *garantie* sont ceux dans lesquels l'objet est toujours

retrouvé avec probabilité 1 lorsque celui-ci est observé par le(s) chercheur(s).

En 2014, Bonato et MacGillivray [6] ont présenté une première généralisation des jeux de policier et voleur qui inclut la majorité de ses variantes. En effet, tous les jeux de poursuite à deux joueurs, à tour de rôle, temps discret, information parfaite et dans lesquels est supposé que les joueurs jouent optimalement, sont inclus dans le modèle de Bonato et MacGillivray. De même, ce modèle inclut tous les jeux de poursuite dits combinatoires (on renvoie au livre de Conway *On Numbers and Games* [15] pour une introduction sur le sujet des jeux combinatoires). Ceux-ci incluent l'ensemble des jeux se jouant à tour de rôle, à information parfaite, sur une structure discrète, tel qu'un graphe, et sans aucun élément de chance. L'exemple typique de jeu combinatoire est le jeu des échecs. Par opposition, le jeu de Stratégo [59], bien qu'il soit totalement déterministe et discret n'est pas considéré comme un jeu combinatoire dû au fait qu'il se joue à information imparfaite. De même, les jeux de chance comme le poker ne sont pas considérés, entre autres de par le manque d'information disponible aux joueurs, mais aussi de par la stochasticité impliquée. Les jeux combinatoires ne comprennent donc aucun élément probabiliste.

Récemment, certains chercheurs tels que Pralat et Kehagias [25] et Komarov et Winkler [27] ont décrit un jeu, le policier et voleur saoul, dans lequel le voleur se promène sur le graphe d'une manière aléatoire : chacun de ses mouvements est décrit par une marche aléatoire uniforme sur les sommets du graphe. Autrement dit, chaque fois que celui-ci veut faire un mouvement, soit sortir d'un sommet, il choisit aléatoirement le prochain sommet parmi l'ensemble des sommets voisins. En général, cette stratégie est sous-optimale. Toutefois, cette description forme la base d'une vaste classe de jeux intéressants supposant chacun que le voleur agisse d'une certaine manière aléatoire. En effet, ce comportement aléatoire ne forme pas seulement l'exemple de base de tout bon livre d'introduction sur les marches aléatoires (le problème de l'ivrogne qui veut rentrer chez lui) [5, 48], le fait de permettre du hasard dans les jeux de policiers et voleurs permet aussi de prendre en considération des situations plus réalistes que les jeux de policiers et voleurs classiques. Un bon exemple de ceci sera étudié au chapitre 4 lors de la résolution du problème dit de l'OSP. Bref, comme ces jeux ne peuvent être décrits par le modèle de Bonato et MacGillivray, il est naturel de songer à étendre ce modèle pour incorporer les jeux comportant des éléments stochastiques.

Ainsi, ce chapitre présente un modèle de jeux de policiers et voleurs qui est plus vaste que celui de Bonato et MacGillivray. L'objectif principal de ce modèle est d'inclure les jeux tels que le policier et voleur saoul où le voleur joue aléatoirement. Étant donné la nature probabiliste de ce jeu, notre modèle ne réutilise pas le même cadre théorique que celui de Bonato et MacGillivray, celui des jeux combinatoires.

3.1.1 Concept de solutions dans les jeux de poursuite

Dans les jeux de policiers et voleurs, on s'intéresse généralement à savoir comment *résoudre* un jeu. Cette question est universelle à la théorie des jeux ; on devrait typiquement définir un *concept de solution* tel que *l'équilibre de Nash* pour s'assurer que cette question soit bien définie. Dans le cas des jeux de policiers et voleurs, on adopte typiquement le point de vue du policier et on ne s'intéresse qu'à la question de savoir s'il est possible, et si oui comment, que celui-ci capture le voleur. Dans notre cadre, cette question se traduit par la détermination de la probabilité de capture du voleur par le policier.

On remarque que la solution de plusieurs jeux de policiers et voleurs prend essentiellement la même forme puisqu'ils peuvent être résolus à l'aide d'une expression récursive logique. En effet, Nowakowski et Winkler [39] en 1983 ont défini un premier préordre $x \preceq y$ signifiant que lorsque le policier se place en y il peut assurément capturer le voleur en x . Cette relation a été étendue 30 ans plus tard par Clarke et MacGillivray [12] pour résoudre les jeux à k -policiers en les définissant sur des produits forts de graphes, c'est-à-dire y est élément de V^k ¹. Il apparaît donc qu'en transformant le graphe sur lequel le jeu se déroule, on peut toujours définir une relation \preceq de manière à résoudre le jeu.

Malgré tout, de telles relations ne peuvent être définies que pour des jeux déterministes, soit des jeux dans lesquels on suppose que les joueurs peuvent jouer optimalement. Dans les jeux à composantes stochastiques, ces relations peuvent être reproduites autrement. En effet, dans un article récemment publié sur le problème de l'OSP [52] est présentée une fonction de récurrence $w_n(x, y)$ donnant la probabilité qu'en n pas ou moins un voleur situé sur le sommet x se fasse capturer par un policier situé en y . Cette relation, définie dans le jeu du policier et voleur saoul, est similaire à la relation $x \preceq_n y$ de Nowakowski et Winkler tout en étant plus générale puisqu'elle permet de modéliser des événements aléatoires. Un aspect intéressant de celle-ci, comme des relations \preceq d'ailleurs, est sa calculabilité en temps polynomial. On peut donc s'interroger à savoir jusqu'où peut-on étendre la définition des jeux de policiers et voleurs pour inclure des éléments aléatoires tout en permettant à ceux-ci d'être résolubles en temps polynomial par une récurrence w_n .

Finalement, comme on peut s'interroger sur le calcul de la stratégie optimale des policiers, on peut aussi s'interroger sur la complexité algorithmique de ce calcul. Kinnersley [26] a démontré une conjecture de Goldstein et Reingold [20] selon laquelle il est EXPTIME-complet de déterminer si le nombre minimal de policiers requis sur un graphe G pour capturer un voleur

1. Le produit fort de G et H est noté $G \boxtimes H = (V = V(G) \times V(H), E \subseteq E(G) \times E(H))$. Il a pour ensemble de sommets V le produit cartésien des ensembles de sommets de G et de H . Deux paires de sommets $(u, u') \in V(G), (v, v') \in V(H)$ sont adjacentes si et seulement si

- $u \in N_G[v]$ et $u' = v'$ ou
- $u = v$ et $u' \in N_H[v']$ ou
- $u \in N_G[v]$ et $u' \in N_H[v']$.

On écrit G^k pour le graphe G multiplié k fois avec lui-même sous le produit fort.

est plus petit qu'un entier k lorsque G et k sont donnés en entrée. Autrement dit, la question de minimiser le nombre de policiers à envoyer en patrouille est EXPTIME-complet, et ce même pour un jeu de policiers et voleurs très simple. D'un autre côté, une variante du jeu de policier et voleur dans laquelle le voleur est invisible, le *Optimal Search Path problem* (voir le chapitre 4), est NP-complet. De plus, l'algorithme présenté par Clarke et MacGillivray [12] pour résoudre le jeu de policiers et voleurs à k policiers nécessite $O(|V(G)|^{2k+2})$ opérations. Aucun algorithme connu n'est plus rapide que celui-ci. Bref, la complexité algorithmique des questions reliées aux problèmes de policiers et voleurs semble généralement élevée. Il est donc important de prendre cet aspect en considération dans notre étude.

De nombreux résultats sur les jeux stochastiques (présentés dans le chapitre 2) soutiennent que des problèmes similaires à ceux présentés au paragraphe précédent, tel que déterminer si un jeu possède un équilibre de Nash, sont NP-complet, sans plus. Ceci semble un peu surprenant compte tenu du fait que cette même question devrait être EXPTIME-complète dans les jeux de policiers et voleurs. Ceci s'explique par le fait que les jeux stochastiques se décrivent à partir d'ensemble d'états S généraux. Les résultats de complexité sont donc construits à partir de $|S|$, obscurcissant le fait que $|S|$ peut être très élevée. Notamment, dans le jeu à k -policiers on doit avoir $|S| = O(|V(G)|^k)$, qui est de taille exponentielle selon le nombre de policiers. On devra donc faire attention lorsque viendra le temps, à la section 3.3, de présenter des résultats de complexité algorithmique qui dépendront de $|S|$ pour obtenir des résultats significatifs.

Ce chapitre est organisé comme suit. À la section 3.2 est présenté notre modèle de jeux de policiers et voleurs. Cette même section présente aussi la récursion w_n qui permet de résoudre ce modèle. La section suivante, section 3.3, présente des résultats de complexité algorithmique, notamment sur la complexité de la relation w_n . Des résultats sur la stationnarité de cette relation sont présentés à la section 3.4. Comme les jeux de policiers et voleurs sont généralement joués sur des graphes, on présente un modèle de jeux de policiers et voleurs se déroulant aussi sur un graphe à la section 3.5. Finalement, la section 3.6 présente des exemples d'applications du modèle, et la conclusion suit à la section 3.7.

3.2 Un jeu de poursuite abstrait à information parfaite

La prochaine définition présente un modèle général pour décrire les jeux de policiers et voleurs à information parfaite. Le modèle est formulé en deux parties : dans la définition 3.2.1 sont présentées les règles du jeu suivies de quelques exemples de jeux de policiers et voleurs encodés dans notre modèle.

Dans le reste du texte on utilise fréquemment les lettres c pour *cop* et r pour *robber* pour faciliter la lecture et puisque cette notation est très utilisée dans la littérature.

Définition 3.2.1. *Un jeu de poursuite discret, dit jeu de policiers et voleurs, \mathcal{G} est joué par*

deux joueurs, une équipe de policiers et une équipe de voleurs. Les policiers bougent en premier et tentent de capturer les voleurs. \mathcal{G} se déroule à information parfaite, à tour de rôle, sur une structure discrète et se décrit par le tuple suivant

$$\mathcal{G} := (S, I, F, A, T_{\text{cop}}, T_{\text{rob}}). \quad (3.1)$$

De plus, les conditions suivantes doivent être respectées

1. $S = S_{\text{cop}} \times S_{\text{rob}} \times S_o$ est un ensemble fini d'états représentant les configurations possibles du jeu. S_{cop} et S_{rob} sont les positions possible des policiers et des voleurs alors que S_o est un ensemble pouvant contenir d'autres informations nécessaires.
2. $I \subset S$ est l'ensemble des états initiaux du jeu.
3. $F \subset S$ est l'ensemble des états gagnants (finaux) du jeu pour les policiers.
4. $A = A_{\text{cop}} \cup A_{\text{rob}}$ est l'ensemble fini des actions des policiers et des voleurs.
5. $T_{\text{cop}} : S \times A_{\text{cop}} \times S \rightarrow [0, 1]$ est une fonction satisfaisant $\sum_{s' \in S} T_{\text{cop}}(s, a, s') \in \{0, 1\}$ pour tous $s \in S$ et $a \in A_{\text{cop}}$. L'action a peut être jouée à partir de l'état s si et seulement si cette valeur est 1. Cette fonction est spécifique aux policiers.
6. $T_{\text{rob}}(s, a, s')$ est une fonction similaire à $T_{\text{cop}}(s, a, s')$, mais spécifique aux voleurs.

Pour un état fixé $s \in S$, on écrit $A_x(s) = \{a \in A_x(s) \mid T_x(s, a, S) = 1\}$ l'ensemble des actions jouables à cet état, pour le joueur $x \in \{\text{cop}, \text{rob}\}$. On peut extraire de la fonction T_{cop} la fonction $T_{\text{cop}}(s, a) : \mathcal{P}(S) \rightarrow [0, 1]$ interprétée comme une distribution sur S ou la fonction constante 0 et de même pour T_{rob} ². C'est souvent cette forme qui sera privilégiée lors des définitions. On écrit $T_{\text{cop}}(s, a, S')$ plutôt que $T_{\text{cop}}(s, a)(S')$ pour $S' \subseteq S$. On utilise aussi parfois la notation $s_x, x \in \{\text{cop}, \text{rob}, o\}$ pour dénoter la projection d'un état $s \in S$ sur l'ensemble S_x . L'ensemble S_o n'est pas souvent utilisé, mais le sera dans l'exemple 3.6.5 sur les graphes dynamiques dont la structure peut varier au fil du temps. Pour simplifier certains résultats, tel que le théorème 3.2.15, on suppose que I ne contient qu'un seul élément. Ceci se fait sans trop de perte de généralité comme on le verra dans les exemples qui viennent.

Dans ce qui suit, on note Dist_B l'ensemble des distributions discrètes sur un ensemble B et $\mathcal{U}_B \in \text{Dist}_B$ la distribution uniforme discrète sur le même ensemble. Une distribution δ sur S est Dirac s'il n'existe qu'un seul élément $s \in S$ tel que $\delta(s) > 0$, auquel cas $\delta(s) = 1$. On écrit δ_s la distribution Dirac sur un élément s .

On dit qu'un jeu est déterministe lorsque les distributions définies par T_{cop} et T_{rob} sont concentrées en un seul point, autrement dit, si $T_{\text{cop}}(s, a)$ et $T_{\text{rob}}(s, a)$ sont Dirac pour tout $s \in S$ et $a \in A$.

2. Formellement, pour tout $s \in S, a \in A$, $T_{\text{cop}}(s, a)$ satisfait $T_{\text{cop}}(s, a)(\emptyset) = 0$, $T_{\text{cop}}(s, a)(S) = 1$ et pour tout $(B_i)_{i \in \mathbb{N}}, B_i \subseteq S$ une famille d'ensembles disjoints on a $\sum_i T_{\text{cop}}(s, a)(B_i) = T_{\text{cop}}(s, a)(\bigcup_{i \in \mathbb{N}} B_i)$. De même pour $T_{\text{rob}}(s, a)$.

Définition 3.2.2. Soit \mathcal{G} un jeu caractérisé par la définition 3.2.1. Une partie \mathcal{H} de \mathcal{G} est une suite possiblement infinie $(SA_{\text{cop}}SA_{\text{rob}})^\omega$ d'états et d'actions de \mathcal{G} . Elle s'arrête lorsqu'un état final $f \in F$ est atteint. Ainsi, \mathcal{H} est fini si et seulement si son dernier élément fait partie de F , dans ce cas les policiers ont gagné et les voleurs ont perdu. Si \mathcal{H} est infini, les voleurs ont gagné. On suppose toujours que S est non-vide, de même que $A_x(s)$ pour tout $s \in S$ et $x \in \{\text{cop}, \text{rob}\}$.

Une partie se déroule donc de la manière suivante : partant d'un état s , les policiers choisissent une action $a_{\text{cop}} \in A_{\text{cop}}(s)$ à partir de laquelle un nouvel état s' est choisi aléatoirement selon $T_{\text{cop}}(s, a_c)$, enfin les voleurs jouent une action $a_{\text{rob}} \in A_{\text{rob}}(s')$ à partir de laquelle le prochain état s'' est pigé avec probabilité $T_{\text{rob}}(s', a_{\text{rob}}, s'')$. Cette suite d'états et d'actions décrit un tour de jeu. Le jeu se termine si et seulement si s' ou s'' est élément de F .

Pour préciser que le jeu se déroule sur une structure discrète, on requiert que S soit un ensemble fini. On aurait pu obtenir un jeu discret en considérant S comme dénombrable, mais certains des résultats plus loin sont simplifiés lorsque S est fini. Toute information pertinente qui n'est pas incluse dans S_{cop} ou S_{rob} , par exemple dans certains jeux où le compteur de tours est important, se retrouve dans S_o . Dans ce qui suit, S_o ne sera pas fréquemment utilisé puisque peu d'exemples nécessitent cette composante, on écrira alors simplement $S = S_{\text{cop}} \times S_{\text{rob}}$. Généralement, les jeux se déroulent sur des graphes, auquel cas S_{cop} et S_{rob} sont simplement des ensembles de sommets. Les policiers jouent des actions choisies dans $A_{\text{cop}}(s)$ et transitent vers des états selon la fonction de transition T_{cop} , alors que les voleurs choisissent des actions dans les ensembles $A_{\text{rob}}(s)$ et transitent vers d'autres états selon la fonction T_{rob} .

La définition 3.2.1 est illustrée sur un premier exemple, le jeu classique de policier et voleur.

Exemple 3.2.3 (Jeu du policier et voleur classique). Soit $G = (V, E)$ un graphe fini. On pose $i_{\text{cop}}, i_{\text{rob}} \notin V$ deux éléments qui servent de points de départ pour le policier et le voleur. Dans ce jeu, les joueurs se déplacent sur les sommets du graphe, en choisissant leurs successeurs parmi leurs voisins. Les états finaux sont ceux pour lesquels le policier partage la même position que le voleur. Les différents ensembles sont :

$$\begin{aligned} I &:= \{(i_{\text{cop}}, i_{\text{rob}})\} \\ S &:= I \cup V \times V \\ F &:= \{(x, x) \in V^2\} \\ A_{\text{cop}} &:= V \\ A_{\text{rob}} &:= V \end{aligned}$$

Soit $(c, r) \in S$. Alors, soit $c' \in A_{\text{cop}}$ et $r \in A_{\text{rob}}$. On pose :

$$T_{\text{cop}}((c, r), c', (x, r)) := \begin{cases} 1, & \text{si } x = c' \text{ et } (c = i_{\text{cop}} \text{ ou } c' \in N[c]), \\ 0, & \text{sinon.} \end{cases}$$

$$T_{\text{rob}}((c, r), r', (c, x)) := \begin{cases} 1, & \text{si } x = r' \text{ et } (r = i_{\text{rob}} \text{ ou } r' \in N[r]), \\ 0, & \text{sinon.} \end{cases}$$

Selon la définition 3.2.2, le voleur joue en second. Il n'est donc pas nécessaire de préciser la condition $c \neq i_{\text{cop}}$ dans la fonction T_{rob} . Il n'est pas non plus nécessaire d'écrire explicitement le cas $c = r$ puisque dans ce cas la partie est terminée. Les fonctions T_{cop} et T_{rob} encodent les ensembles d'actions valides pour chaque joueur. De ces fonctions, on peut déduire les ensembles d'actions à partir d'états particuliers. Par exemple, lorsque le policier se trouve sur un état $(c, r) \in S \setminus I$, $A_{\text{cop}}(c, r)$ est simplement $N[c]$. De même, partant du même état, le voleur choisit une action parmi $A_{\text{rob}}(c, r) = N[r]$.

La stochasticité du jeu de la définition 3.2.1 est motivée par l'exemple suivant dit du policier et voleur saoul. Celui-ci se décrit de manière très similaire au jeu précédent excepté que le voleur se déplace maintenant de manière aléatoire sur le graphe.

Exemple 3.2.4 (Jeu du policier et voleur saoul). Seule la fonction de transition du voleur change, le reste de la structure est invariant. Le voleur, après avoir quitté sa position initiale, se déplace uniformément aléatoirement sur l'ensemble de ses voisins, ce qui revient à ignorer son action $r' \in A_{\text{rob}}$. Soit $(c, r) \in S$ et $r' \in A_{\text{rob}}$, alors T_{rob} est :

$$T_{\text{rob}}((c, r), r') := \begin{cases} \delta_{r'}, & \text{si } r = i_{\text{rob}}, \\ \mathcal{U}_{N[r]}, & \text{autrement.} \end{cases}$$

On pourrait de manière équivalente poser $A_{\text{rob}}(s) = \{1\}$ lorsque $s \notin S \setminus I$, puisque l'action du voleur est ignorée. Une variante de la fonction de transition T_{rob} consisterait à poser que le voleur a une probabilité positive d'aboutir à $r' \in A_{\text{rob}}(c, r)$ qui dépendrait de la distance entre r et r' .

Les graphes de la figure 3.1 illustrent un exemple de mouvements des joueurs dans un jeu de policier et voleur saoul. Ici, l'ensemble d'états S contient les paires de positions policier et voleur $(C_0, R_0), (C_1, R_0), (C_1, R_1)$. Les sommets en rouges représentent les positions des joueurs alors que ceux en rose représentent les actions possibles du voleur. Les trois graphes représentent donc trois états consécutifs dans une partie. Le graphe \mathcal{G}_1 présente l'état courant suivi du graphe \mathcal{G}_2 résultant de l'action du policier. Le graphe \mathcal{G}_3 résulte de la réplique du voleur.

Dans notre article récemment présenté à CP 2015 [52] est décrite une variante du jeu précédent, le policier contre le voleur saoul et *évasif*, dans lequel le voleur a une probabilité positive de

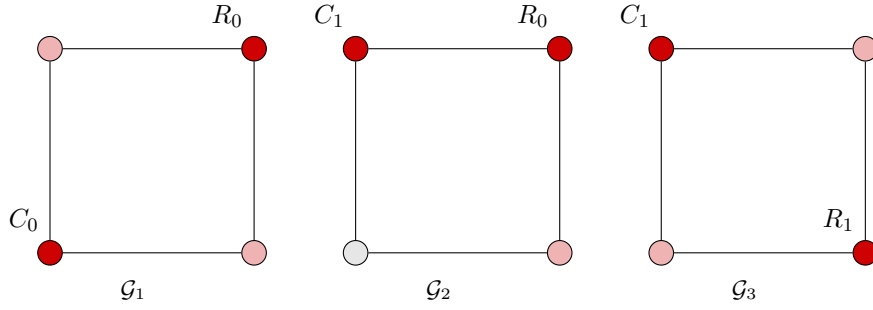


FIGURE 3.1 – Exemple de deux mouvements des joueurs dans le jeu de policier et voleur saoul.

ne pas se faire capturer lorsque le policier et lui se croisent sur un sommet. Ce jeu motive partiellement la généralité de la fonction de transition T_{rob} qui permet d'encoder différents événements aléatoires. La principale différence entre ce jeu et le précédent est l'incapacité du policier à garantir la capture du voleur lorsque ceux-ci se croisent. Autrement dit, le voleur a une probabilité d'échapper au policier lorsque les deux se retrouvent sur le même sommet.

Exemple 3.2.5 (Jeu du policier contre le voleur saoul et évasif). La structure du jeu est encore une fois très similaire à celle du jeu de policier et voleur classique. On ajoute cependant au jeu une case particulière j^* , une prison. Les états initiaux sont les mêmes que précédemment. On écrit donc $S = I \cup V \times V \cup \{(j^*, j^*)\}$. Les états finaux sont ceux dans lesquels les deux joueurs sont en prison, $F = \{(j^*, j^*)\}$.

Les actions restent inchangées par rapport aux deux jeux précédents. On doit cependant réécrire les fonctions de transition. Lorsque les deux joueurs ne se rencontrent pas, ils se déplacent sur G de la même manière que dans le jeu du policier et voleur saoul. Cependant, lorsque le policier rencontre le voleur sur le sommet r , il a une certaine probabilité $p(r)$ de l'attraper, où la fonction p satisfait $p : V_r \rightarrow [0, 1]$ et où $V_r \subseteq V$ est l'ensemble des sommets accessibles au voleur. Il est possible que cette probabilité ne soit pas la même sur tous les sommets, c'est pourquoi p est une fonction de r . La fonction de transition est donc de la forme :

$$T_{\text{rob}}((c, r), r') := \begin{cases} \delta_{r'}, & \text{si } r = i_{\text{rob}} \\ \mathcal{U}_{N[r]}, & \text{si } c \neq r \text{ et } r \neq i_{\text{rob}} \\ D_r, & \text{si } c = r \text{ et } r \neq i_{\text{rob}}, \end{cases}$$

où

$$D_r(x) := \begin{cases} \frac{(1-p(r))}{|N[r]|}, & \text{si } x \in N[r], \\ p(r), & \text{si } x = (j^*, j^*). \end{cases}$$

La fonction D_r est invoquée quand le policier et le voleur se croisent sur un sommet, c'est-à-dire quand $c = r$. Cela se produit *pendant* ou à la fin du tour du policier, car cette égalité

est vérifiée avant que le voleur ne choisisse son action r' . Celle-ci n'est exécutée comme voulu que si le policier ne l'a pas attrapé, avec probabilité $1 - p(r)$: l'état d'arrivée est alors choisi uniformément dans le voisinage de r . La fonction T_{rob} peut être définie avec des distributions plus complexes que la distribution uniforme, il suffit de remplacer \mathcal{U}_B par un élément de Dist_B dans la fonction précédente pour chaque ensemble B . De même, dans D_r on pourrait remplacer le facteur $\frac{1}{|N[r]|}$ par une distribution quelconque de $\text{Dist}_{N[r]}$.

Finalement, on présente le jeu du voleur rapide avec zone de surveillance tel qu'énoncé dans le mémoire de Marcoux [33]. Cet exemple est repris à la section 3.5 où la notation sera modifiée pour cadrer avec la définition de cette section.

Exemple 3.2.6 (Jeu du policier et voleur rapide). Ce jeu est similaire au jeu classique excepté pour deux points : le voleur peut se déplacer le long de chemins de longueur finie, plutôt que seulement le long d'une arête, et, en contrepartie, le policier peut capturer son adversaire lorsqu'il se trouve dans son voisinage. Plus généralement, le policier pourrait capturer le voleur lorsqu'il se trouve dans un certain ensemble de sommets contrôlés par le policier, que l'on nomme zone de surveillance. Cette zone de surveillance peut simuler l'utilisation d'une arme par le policier. Les états doivent donc être légèrement modifiés. Ceux-ci sont maintenant des triplets contenant, en plus des positions des deux joueurs, l'ensemble de sommets surveillés par le policier, en deuxième position. À l'état initial, la zone de surveillance du policier est vide, le voleur ne peut pas se faire attraper avant son premier coup. On utilise encore une fois un état prison $j^* \notin V$. Lorsque les deux joueurs s'y trouvent, le jeu se termine et le voleur a perdu. Ainsi, on pose :

$$\begin{aligned} I &:= \{(i_{\text{cop}}, \emptyset, i_{\text{rob}}) \mid i_{\text{cop}}, i_{\text{rob}} \notin V\}, \\ F &:= \{(j^*, \emptyset, j^*)\}, \\ S &:= I \cup \{(c, N[c], r) \mid c, r \in V\} \cup F \end{aligned}$$

l'ensemble des états de jeu.

La fonction de transition du policier est simple. Elle encode sa zone de surveillance qui est déterminée après son action. Soit $(c, N[c], r) \in S$ l'état courant et $c' \in N[c]$ l'action du policier. Le policier choisit un sommet lorsqu'il se trouve à l'état initial, autrement il se déplace dans le voisinage de sa position courante. Pour simplifier le problème, on a considéré que le policier n'avait pas le choix de sa zone de surveillance. Dans une version plus générale, il pourrait choisir un voisin vers lequel se déplacer ainsi qu'une zone de surveillance. On écrit donc :

$$T_{\text{cop}}((c, N[c], r), c') = \begin{cases} \delta_{(c', N[c'], r)} & \text{si } (c = i_{\text{cop}} \text{ et } c' \in V) \text{ ou si } (c \in V \text{ et } c' \in N[c]), \\ 0, & \text{sinon.} \end{cases}$$

À son tour, lorsqu'il se tient sur un sommet $r \in V$, l'action du voleur consiste à choisir un chemin de longueur finie. Soit $(c, N[c], r) \in S$ et $\pi = (r = r_1, r_2, \dots, r_n) \in \mathcal{P}_r$, $n > 0$ un

chemin. La fonction de transition des voleurs est :

$$T_{\text{rob}}((c, N[c], r), \pi) := \begin{cases} \delta_{(c, N[c], r_n)} & \text{si } r = i_{\text{rob}} \text{ et } r_n \in V \setminus N[c], \\ \delta_{(c, N[c], r_n)} & \text{si } r \in V \text{ et } r_i \notin N[c], \forall 1 \leq i \leq n, \\ \delta_{(j^*, \emptyset, j^*)} & \text{sinon.} \end{cases}$$

Le voleur est donc assuré d'atteindre sa destination r_n pourvu qu'il ne traverse jamais la zone de surveillance du policier sur son chemin π . Si ceci se produit, alors le voleur est conduit à la prison (j^*, \emptyset, j^*) . Plus loin dans ce texte, à la section 3.5, est présenté ce même jeu, mais cette fois avec la possibilité que le voleur échappe aux tentatives de capture du policier. En effet, l'exemple 3.5.2 décrit un jeu dans lequel les distributions internes à la fonction T_{rob} ne sont pas Dirac. Ce jeu est alors écrit avec la notation de la définition 3.5.1 qui utilise plutôt un ensemble d'arêtes comme zone de surveillance. Ici, il était plus naturel d'utiliser un ensemble de sommets et l'écriture en est simplifiée.

Ainsi, à cause de sa définition assez large, il est possible d'encoder une grande variété d'événements aléatoires résultant de l'action des policiers ou des voleurs. Autrement dit, plusieurs distributions sur l'ensemble d'états conditionnelles à l'état courant ainsi qu'à une action des policiers ou des voleurs peuvent être encodées dans notre modèle.

Le prochain exemple illustre comment notre modèle intègre les événements stochastiques en incluant naturellement un processus stochastique très connu dans la littérature, la chaîne de Markov.

Exemple 3.2.7 (Chaîne de Markov finie). La chaîne de Markov sur l'espace d'états S fini peut être modélisée comme un jeu de policiers et voleurs avec la définition 3.2.1. Soit $i_{\text{cop}}, i_{\text{rob}} \notin S$ les positions de départ des joueurs, alors $I = \{(i_{\text{cop}}, i_{\text{rob}})\}$. Il n'y a aucun état final, $F = \emptyset$. L'ensemble d'actions ne contient qu'un seul élément, $A = \{1\}$. Soit M la matrice de transition de la chaîne de Markov que l'on simule et q sa distribution initiale. Alors, les fonctions de transitions sont :

$$T_{\text{cop}}(s, 1, s') = T_{\text{rob}}(s, 1, s') := \begin{cases} q(s'), & \text{si } s \in I, \\ M(s, s'), & \text{sinon.} \end{cases}$$

Ainsi, les actions des joueurs n'ont aucune influence sur le déroulement du jeu. De plus, les probabilités de transition sont égales pour les deux joueurs ce qui simule le fait qu'un seul agent fait partie du système. Le jeu évolue donc de manière purement aléatoire.

L'exemple précédent est trivial, un exemple plus intéressant est le suivant.

Exemple 3.2.8 (Processus de Markov décisionnel). Le processus de Markov décisionnel (*Markov decision process*, MDP) est constitué d'une chaîne de Markov sur un espace d'état S_{MDP} et

d'un joueur qui à chaque état s doit choisir une action a dans un ensemble A_{MDP} . Le système évolue selon la probabilité de transition $\mathbb{P}[s' | s, a]$ qui est une généralisation de la probabilité de transition classique de la chaîne de Markov puisqu'elle dépend d'une deuxième variable, l'action a . Le joueur doit tenter de contrôler le système en choisissant l'action a de telle sorte à atteindre l'état s' maximisant son gain. On se restreint aux MDPs dans lesquels le gain est 1 lorsqu'un état final est atteint. Les jeux de policiers et voleurs, contrairement aux MDP, n'utilisent pas explicitement la notion de gain et c'est pourquoi on se limite à un gain simple. Supposons que S_{MDP} et A_{MDP} soient finis. On note aussi $\mathbb{P}[s]$ pour la probabilité que l'agent se trouve initialement à l'état $s \in S_{\text{MDP}}$. Pour le jeu de policiers et voleurs, on suppose que les joueurs agissent dans l'ensemble S_{MDP} et débutent sur un état initial $(i_{\text{cop}}, i_{\text{rob}}) \notin S_{\text{MDP}}$. De plus, $I = \{(i_{\text{cop}}, i_{\text{rob}})\}$ et il n'y a aucun état final, c'est-à-dire $F = \emptyset$. L'ensemble d'actions des joueurs est A_{MDP} .

Les fonctions de transition sont :

$$T_{\text{cop}}(s, a, s') = T_{\text{rob}}(s, a, s') := \begin{cases} \mathbb{P}[s'] & \text{si } s \in I \\ \mathbb{P}[s' | a, s] & \text{sinon.} \end{cases}$$

Les deux joueurs, policiers et voleurs, simulent le rôle d'un seul joueur agissant dans un MDP.

3.2.1 Stratégies

On définit ici le concept de *stratégie* qui sera utilisé un peu plus bas pour écrire le théorème principal de ce chapitre. Celui-ci définit la stratégie optimale du jeu \mathcal{G} pour les policiers. On ne présente ici que les stratégies sans mémoire et déterministes puisque ce sont les seules stratégies nécessaires à la compréhension du reste de ce chapitre. Les définitions sont écrites sous la forme générale et ne sont pas directement liées au jeu défini par la définition 3.2.1.

Une *stratégie* est généralement décrite simplement comme un plan d'action à suivre qu'utilisent les joueurs dans un jeu pour maximiser leurs gains ou leurs chances de gagner. On utilise ici le concept de stratégies dites *sans mémoire*, qui sont des stratégies qui ne dépendent que de l'état de jeu courant. À l'opposé, dans certains jeux il est intéressant pour les joueurs d'analyser l'ensemble des actions qui ont été jouées auparavant avant de prendre une décision.

Définition 3.2.9. *Une stratégie pure ou déterministe et sans mémoire est une fonction $\sigma : S \rightarrow A$ prescrivant à un joueur une action à choisir à chaque état de jeu possible.*

On dit d'une stratégie pure qu'elle est *déterministe* puisqu'elle est une fonction, alors que le terme *sans mémoire* renvoie au fait qu'elle ne dépend que d'un seul état et non de l'historique de jeu complet. Dans la littérature, il est démontré que sous certaines conditions l'utilisation de stratégies sans mémoire se fait sans perte de généralité. On renvoie à l'annexe A, et plus particulièrement au théorème A.1.4, pour plus de détails. Notamment, on déduit d'un des

résultats du théorème A.1.4 que l'utilisation de stratégies sans mémoire se fait sans perte de généralité.

3.2.2 Conditions de victoire dans les jeux de policiers et voleurs

Dans cette section sont définis les objectifs d'un jeu de poursuite \mathcal{G} construit selon la définition 3.2.1 qui serviront d'objectifs à maximiser pour chaque joueur.

Dans ce qui suit, on suppose que les joueurs peuvent être limités à jouer au plus $N \in \mathbb{N} \cup \{\infty\}$ tours, quoique N peut ne pas être borné. On écrit aussi Ω_c et Ω_r pour les ensembles de stratégies que les joueurs peuvent utiliser.

On définit la *valeur* d'un jeu de poursuite \mathcal{G} comme la probabilité qu'en N tours les voleurs soient capturés lorsque les deux joueurs jouent optimalement. En théorie des jeux, les joueurs cherchent typiquement à maximiser des gains (par exemple des profits) et la valeur d'un jeu représente essentiellement les gains que chaque joueur peut espérer obtenir en suivant sa stratégie optimale [56, 51]. Soit $T_{\sigma_{\text{cop}}, \sigma_{\text{rob}}}$ la variable aléatoire qui, étant donné deux stratégies σ_{cop} et σ_{rob} , retourne le nombre de tours avant la capture, s'il y a capture, et retourne l'infini sinon. Dans un jeu à deux joueurs, la valeur du jeu pour un joueur est sa probabilité de gagner, alors que la valeur du jeu est la probabilité que le premier joueur gagne en jouant optimalement contre un adversaire intelligent [56]. Alors, la valeur du jeu est la probabilité p_N^* qu'il y ait une capture en au plus N tours lorsque les joueurs suivent leurs stratégies optimales, c'est-à-dire :

$$p_N^* := \max_{\sigma_c \in \Omega_c} \min_{\sigma_r \in \Omega_r} \mathbb{P} [T_{\sigma_{\text{cop}}, \sigma_{\text{rob}}} \leq N]. \quad (3.2)$$

Selon la théorie des jeux stochastiques simples [56], la valeur optimale p_N^* est atteinte avec des stratégies sans mémoire et déterministes et ainsi on ne perd aucune généralité en définissant l'équation (3.2) avec des ensembles Ω_c, Ω_r de telles stratégies. On renvoie à la section A.1 de l'annexe A pour une description des jeux stochastiques ainsi que de certains résultats intéressants utilisés ici.

L'objectif des joueurs est donc pour les policiers de maximiser la probabilité de capturer les voleurs dans un temps fini N , alors que les voleurs cherchent l'opposé. L'équation (3.2) retourne toujours 0 ou 1 dans des jeux déterministes tels que les jeux de policiers et voleurs décrits par le modèle de Bonato et MacGillivray [6]. Cependant, nous voulons étudier des jeux qui sont stochastiques, menant à des valeurs optimales prenant n'importe quelle valeur p_N^* dans l'intervalle $[0, 1]$. Finalement, comme dans les jeux de policiers et voleurs on s'intéresse à des graphes *policiers-gagnants*, on définit certains types de conditions de victoire dans les jeux généraux.

Définition 3.2.10. *Soit \mathcal{G} un jeu de poursuite et Ω_c, Ω_r deux ensembles de stratégies sans mémoires et déterministes. On dit que \mathcal{G} est $c(1, N)$ -gagnant si et seulement si les policiers*

ont probabilité 1 de gagner en temps N , soit

$$p_N^* = 1.$$

Plus généralement, on parle de jeu $c(p, N)$ -gagnant si les policiers peuvent s'assurer une victoire avec une probabilité au moins p en N tours,

$$p_N^* \geq p.$$

Finalement, on dit qu'un jeu est gagnant à la limite si les policiers peuvent gagner en temps non borné,

$$p_\infty^* = 1,$$

Où $p_\infty^* := \lim_{n \rightarrow \infty} p_n^*$.

La première condition est la plus forte à cause de la limite de temps, mais la deuxième est plus réaliste. La troisième, quant à elle, est un cas dégénéré. En effet, si la probabilité de transition des voleurs T_{rob} est telle que $T_{\text{rob}}(s, a, s') > 0$ pour toute paire d'états s et s' et action $a \in A_{\text{rob}}$, alors les policiers sont assurés de gagner en temps non borné puisque les voleurs vont toujours avoir une probabilité positive d'entrer (et de rester) dans un état final. Cependant, la troisième condition est plus souvent utilisée dans les jeux de policiers voleurs puisque les chercheurs travaillant dans ce domaine s'intéressent rarement aux jeux qui ont des limites de temps. Suivant la définition 3.2.10 et le modèle de Bonato et MacGillivray, on définit qu'un jeu est *policiers-gagnants* si les policiers peuvent gagner en temps non borné contre des voleurs jouant optimalement.

Définition 3.2.11. *Un jeu déterministe \mathcal{G} est dit policiers-gagnant si et seulement si les policiers peuvent gagner en temps non borné contre des voleurs jouant optimalement.*

$$p_\infty^* = 1.$$

Avant de clore cette section, on définit un paramètre de jeu à l'aide de la valeur p_N^* . Dans les jeux de policiers et voleurs usuels, le nombre $c(G)$, dit *cop number*, est souvent l'objet d'étude principal [8]. Celui-ci est le nombre minimal de policiers nécessaires à la capture d'un voleur sur le graphe G dans le jeu de policiers et voleurs classique. Plusieurs généralisations de ce nombre, que l'on considère comme un paramètre de graphe, ont été proposées en suivant les différentes variantes de jeux de policiers et voleurs. D'ailleurs, une des avenues de recherche les plus importantes dans les jeux de policiers et voleurs est la conjecture de Meyniel [8] selon laquelle $c(G) \in O\left(\sqrt{|V(G)|}\right)$. Autrement dit, le nombre de policiers nécessaires à la capture du voleur croît asymptotiquement en fonction de la racine du nombre de sommets d'un graphe. On propose finalement une extension de ce paramètre à des jeux plus généraux.

Définition 3.2.12. *Le (p, n) -cop number $c_p^n(\mathcal{G})$ d'un jeu \mathcal{G} est le nombre minimal de policiers nécessaires à la capture des voleurs avec probabilité au moins p en au plus n tours. Autrement dit, c'est le nombre minimal de policiers nécessaires à ce qu'un jeu \mathcal{G} soit $c(p, n)$ -gagnant.*

Lorsque le nombre de tours n'est pas précisé, on suppose que celui-ci n'est pas borné et on parle simplement du p -cop number. Autrement dit, on pose $c_p(\mathcal{G}) := c_p^\infty(\mathcal{G})$.

3.2.3 Résolution des jeux de policiers et voleurs

Similairement au modèle de Bonato et MacGillivray, on définit une méthode pour résoudre les jeux de poursuite, soit de déterminer la probabilité maximale que les policiers capturent les voleurs, peu importe la stratégie de ces derniers. Cette méthode prend la forme d'une récursion qui associe à un état s la probabilité que celui-ci mène à une capture en au plus n tours.

Définition 3.2.13. *Soit \mathcal{G} un jeu de poursuite défini selon la définition 3.2.1. Supposons que les deux joueurs jouent optimalement et posons :*

$$w_0(s) := \begin{cases} 1, & \text{si } s \in F, \\ 0, & \text{sinon.} \end{cases}$$

$$w_n(s) = \begin{cases} 1, & \text{si } s \in F, \\ \max_{a_c \in A_{\text{cop}}(s)} \sum_{s' \in S} T_{\text{cop}}(s, a_{\text{cop}}, s') \min_{a_r \in A_{\text{rob}}(s')} \sum_{s'' \in S} T_{\text{rob}}(s', a_{\text{rob}}, s'') w_{n-1}(s''), & \text{autrement.} \end{cases} \quad (3.3)$$

Dans la récurrence $w_n(s)$ de la définition précédente, la lettre w fait référence au terme anglais *winning*. Dans les jeux de policiers et voleurs, on cherche généralement des stratégies gagnantes pour les policiers (en anglais *winning strategies*) et c'est implicitement ce que la récurrence calcule.

Remarque 3.2.14. *L'équation (3.3) a une structure relativement simple. On reconnaît les deux opérateurs \max et \min , ce qui est attendu puisque les joueurs cherchent respectivement à maximiser et à minimiser la probabilité de capture et que les policiers jouent en premier. Les deux autres termes, les sommations, sont interprétées comme deux probabilités ou deux espérances. La seconde sommation est interprétée comme la probabilité que l'action des voleurs mène à une capture. De même, la première sommation est perçue comme la probabilité que l'action des policiers leur permette de capturer les voleurs. L'équation complète est donc la probabilité espérée de capture des voleurs lorsque les deux joueurs jouent optimalement.*

L'intérêt principal de la définition 3.2.13 est le théorème suivant, le théorème principal de cette section, qui justifie que $w_n(s)$ est bel et bien la probabilité de capture des voleurs.

Théorème 3.2.15. Soit $w_n(s)$ la récursion définie selon l'équation (3.3) dans un jeu de policiers et voleurs \mathcal{G} . Alors, $w_n(s)$ donne la probabilité que les voleurs se fassent capturer en n tours ou moins,

$$w_n(i) = p_n^*. \quad (3.4)$$

En d'autres mots, \mathcal{G} est $c(p, n)$ -gagnant si et seulement si $w_n(i) \geq p$.

Démonstration. Démontrons ce résultat par induction sur n .

- $n = 0$: Lorsque $n = 0$, les joueurs n'ont aucune action à jouer. Si les voleurs ont terminé leur tour dans un état final, c'est-à-dire si $s \in F$, alors avec probabilité 1 les policiers ont gagné. Autrement, les policiers ont perdu.
- $n = k + 1$: Supposons que le résultat tienne pour $n \leq k$ et soit s l'état courant. Si cet état est final, alors les voleurs sont attrapés en n tours ou moins avec probabilité 1, donc $w_n(s) = 1$ comme voulu. Autrement, laissons les policiers, jouant en premier, choisir une action $a_{\text{cop}} \in A_{\text{cop}}(s)$ après quoi le prochain état s' est pigé selon $T_{\text{cop}}(s, a_{\text{cop}}, s')$. Alors, les voleurs peuvent jouer une action $a_{\text{rob}} \in A_{\text{rob}}(s')$ auquel cas le prochain état s'' sera pigé avec probabilité $T_{\text{rob}}(s', a_{\text{rob}}, s'')$. Par l'hypothèse d'induction, on sait qu'un état final sera visité dans $n - 1$ tours ou moins avec probabilité $w_{n-1}(s'')$ en partant de s'' . Ainsi, la probabilité que les voleurs soient attrapés en n tours ou moins si les policiers ont atteint s' et que les voleurs choisissent l'action a_{rob} est donnée par

$$\sum_{s'' \in S} T_{\text{rob}}(s', a_{\text{rob}}, s'') w_{n-1}(s'').$$

Les voleurs veulent minimiser cette probabilité sur leur ensemble d'actions, ce qui est possible parce que les ensembles S et A_{rob} sont finis. Après que soit jouée l'action a_{cop} , le jeu peut transiter vers n'importe quel état $s' \in S$ et donc avec probabilité

$$\sum_{s' \in S} T_{\text{cop}}(s, a_{\text{cop}}, s') \min_{a_{\text{rob}} \in A_{\text{rob}}(s')} \sum_{s'' \in S} T_{\text{rob}}(s', a_{\text{rob}}, s'') w_{n-1}(s'')$$

les voleurs sont capturés en au plus n tours lorsque les policiers jouent en s . Les policiers veulent maximiser cette quantité et, tel que pour les voleurs, ceci est possible puisque les ensembles considérés sont finis. Ainsi, les policiers doivent jouer l'action

$$\operatorname{argmax}_{a_c \in A_{\text{cop}}(s)} \sum_{s' \in S} T_{\text{cop}}(s, a_c, s') \min_{a_r \in A_{\text{rob}}(s')} \sum_{s'' \in S} T_{\text{rob}}(s', a_r, s'') w_{n-1}(s'').$$

Les choix d'actions à l'état initial donnent donc la probabilité $w_n(i)$. Comme p_n^* est, par définition, la probabilité de capture des voleurs en n tours ou moins, on conclut que $w_n(i) = p_n^*$.

□

À l'aide de la récurrence précédente (3.3) on peut maintenant reformuler le théorème classique des jeux de policiers et voleurs, ce qui est réalisé dans le corollaire suivant. Sous certaines conditions, qui sont étudiées à la section 3.4, la relation w_n devient *stationnaire*. Dans ce cas, il existe un rang $N \in \mathbb{N}$ tel que pour tout $n > N$, pour tout état $s \in S$, on a $w_n(s) = w_{n+1}(s)$. Bref, la probabilité de capture ne change plus à partir d'un certain indice. On laisse tomber l'indice n et on écrit simplement w pour dénoter la partie stationnaire de la relation (3.3).

Corollaire 3.2.16. *Soit \mathcal{G} un jeu de policiers voleurs. Supposons que la relation w_n devienne stationnaire à partir d'un certain rang N , auquel cas on écrit $w = w_N$. Alors, on dit que \mathcal{G} est policiers-gagnant, au sens typique, si et seulement si*

$$w(i) = 1.$$

De plus, le jeu est dit p -policiers-gagnant si et seulement si

$$w(i) \geq p.$$

Finalement, si la relation ne devient jamais stationnaire, alors le jeu est dit gagnant à la limite si et seulement si

$$\lim_{n \rightarrow \infty} w_n(i) = 1.$$

En particulier, pour un certain $N \in \mathbb{N}$ un jeu p -policiers-gagnant est $c(p, n)$ -gagnant pour tout $n \geq N$. Les conditions pour que la relation w_n devienne stationnaire sont présentées à la proposition 3.4.1.

Évidemment, qu'un jeu soit p -policiers-gagnant est une condition moins forte que la condition qu'un jeu soit policiers-gagnants, mais beaucoup plus générale.

3.3 Complexité de calcul dans les jeux de poursuite

Dans cette section sont présentés divers résultats sur la complexité algorithmique reliés au jeu construit selon la définition 3.2.1. Le résultat le plus important de cette section est la proposition 3.3.1 qui donne la complexité algorithmique du calcul de l'équation (3.3) à l'aide d'une approche de programmation dynamique.

3.3.1 La complexité de calcul de la fonction w_n

On ouvre cette section avec un résultat sur la complexité algorithmique du calcul de la fonction w_n définie selon la définition 3.2.13. Cette fonction est calculable par programmation dynamique, mais peut nécessiter beaucoup de calculs, d'autant plus que la complexité est fonction de la taille de l'ensemble d'états. Ainsi, plus le nombre d'états possibles à chaque tour est grand, plus la fonction w_n sera coûteuse. On rappelle aussi qu'un des objectifs premiers de la modélisation des jeux de policiers et voleurs est l'obtention d'une récurrence, l'équation (3.3),

se calculant en temps polynomial. Il est un peu ambigu ici de parler de temps polynomial puisque la définition 3.2.1 utilise des états abstraits et que le nombre d'états $|S|$ peut être exponentiel selon la taille de la structure de jeu. Par exemple, si k policiers se déplacent sur un graphe G , alors $|S| > |V|^k$, ce qui est exponentiel selon k . Bref, le mieux que l'on puisse obtenir ici est une complexité polynomiale selon la taille de S , et on verra que cette complexité est généralement atteinte. On pose $|A_x|$ pour $\max_{s \in S} |A_x(s)|$ où $x \in \{\text{cop}, \text{rob}\}$.

Proposition 3.3.1. *En pire cas et en utilisant une approche de programmation dynamique le calcul de la valeur $w_n(i)$ requiert $O(n|S|^2|A_{\text{cop}}||A_{\text{rob}}|)$ opérations. La complexité spatiale est $O(n|S|)$.*

Démonstration. Soit a_n le nombre d'opérations nécessaires au calcul de la récurrence w_n . Supposons que le calcul des probabilités de transition $T_{\text{cop}}, T_{\text{rob}}$ se font à coût unité (les valeurs pourraient être emmagasinées dans une table). Clairement, $a_0 = 1$. En pire cas, lorsque $n > 0$ tous les éléments des ensembles A_{cop} et A_{rob} doivent être visités pour s'assurer de choisir les actions optimales, cela nécessitant $|A_{\text{cop}}||A_{\text{rob}}|$ opérations. Soit m_c le nombre maximal d'états s' tels que $T_{\text{cop}}(s, a_{\text{cop}}, s') > 0$ sous la somme $\sum_{s' \in S} T_{\text{cop}}(s, a_{\text{cop}}, s')$ pour une action $a_{\text{cop}} \in A_{\text{cop}}(s)$ et m_r définie similairement pour la somme $\sum_{s' \in S} T_{\text{rob}}(s, a_{\text{rob}}, s')$ et pour une action $a_{\text{rob}} \in A_{\text{rob}}(s)$. On a toujours que $m_c, m_r \leq |S|$. Alors, en pire cas,

$$\begin{aligned} a_n &\leq |S|^2|A_{\text{cop}}||A_{\text{rob}}| + a_{n-1} \\ &\leq n|S|^2|A_{\text{cop}}||A_{\text{rob}}| + 1 \end{aligned}$$

où on a supposé que toutes les valeurs de w_{n-1} sont sauvegardées en mémoire. Toutefois, la sauvegarde des valeurs de w_{n-1} nécessite une complexité spatiale de $O(n|S|)$ en pire cas. La complexité finale en temps est donc $O(n|S|^2|A_{\text{cop}}||A_{\text{rob}}|)$. \square

Il apparaît de la proposition précédente que même avec l'utilisation de stratégies *sans mémoire*, qui dépendent seulement de l'état courant, la complexité algorithmique est élevée. On voit donc l'intérêt de ce type de stratégie. Si les stratégies optimales n'avaient pas été sans mémoire, la complexité de calcul aurait été beaucoup plus élevée.

Comme il a été mentionné plus haut, les complexités, autant spatiale que temporelle, dépendent de la taille de trois ensembles abstraits : S , A_{cop} et A_{rob} . Ceci suggère que cette complexité temporelle peut être très élevée puisque le nombre d'actions jouables à chaque état peut être beaucoup plus élevé que le nombre d'états. On peut imaginer que les actions constituent des chemins, auquel cas on pourrait se retrouver avec une complexité exponentielle. D'autant plus que le nombre d'actions admissibles dépend généralement du nombre d'états de jeu possibles, on déduit qu'une condition nécessaire pour que la résolution d'un jeu de policiers et voleurs soit rapide est que le nombre d'actions jouables ne soit pas trop grand par rapport à l'ensemble d'états.

Corollaire 3.3.2. *Calculer la récursion w_n définie par l'équation (3.3) nécessite un nombre polynomial d'opérations en fonction de la taille de l'ensemble d'états S lorsque les ensembles A_{cop} et A_{rob} sont de tailles polynomiales en fonction de la taille de S .*

Démonstration. Ceci est une conséquence directe de la proposition 3.3.1. La complexité de calcul de w_n prend $O(n|S|^2p(|S|)q(|S|))$ lorsque les tailles des ensembles d'action sont polynomiales. \square

Le corollaire 3.3.2 n'empêche pas que le temps de calcul de la fonction w_n soit exponentiel puisque la taille de l'espace d'états pourrait être exponentielle en fonction, par exemple, du nombre de sommets du graphe ou du nombre de policiers. Notamment, l'algorithme de Clarke et MacGillivray [12] utilise le graphe produit G^k et la taille de son ensemble d'états est de l'ordre de $|V(G)|^{2k}$. Cet algorithme requiert donc $O(|V(G)|^{2k+2})$ opérations, ce qui est exponentiel en k .

Le corollaire 3.3.2 est toutefois intéressant puisqu'il présente une condition nécessaire que certains jeux ne respectent pas, du moins a priori. Par exemple, on peut concevoir une variante du jeu de policier et voleur rapide (le voleur est rapide lorsqu'il se déplace sur des chemins) dans laquelle le voleur choisit aléatoirement sa prochaine action. Dans une telle variante, le nombre d'actions permises au voleur est de l'ordre de grandeur du nombre de chemins du graphe, soit d'un nombre exponentiel en fonction du nombre de sommets du graphe. Ce jeu n'est donc pas polynomial a priori. Un cas plus restreint de ce jeu est étudié à la section 3.5 dans lequel on montre que, sous certaines conditions, il n'est pas nécessaire de parcourir l'ensemble des chemins pour trouver l'action optimale du voleur.

3.3.2 Un résultat de difficulté pour les jeux de poursuite

Dans les jeux de policiers et voleurs, on demande souvent de connaître le nombre minimal de policiers nécessaire pour capturer les voleurs. Dans le jeu de k policiers et voleur, il a été démontré par Kinnersley [26] que cette question, dans sa version décisionnelle, est EXPTIME-complète. Autrement dit, cette question est aussi difficile à résoudre que n'importe quelle autre question dans la classe EXPTIME-difficile. Cette classe contient l'ensemble des problèmes de décision résolubles en temps exponentiel sur une machine de Turing déterministe. Comme EXPTIME contient des problèmes qui nécessitent un nombre exponentiel d'opérations à résoudre, les problèmes complets par rapport à cette classe requièrent forcément un nombre exponentiel d'opérations pour être résolus.

Le jeu de k policiers et voleur est une instance particulière s'inscrivant dans notre modèle et est ainsi aussi difficile à résoudre.

Proposition 3.3.3. *La question :*

Étant donné un jeu de poursuite \mathcal{G} décrit selon la définition 3.2.1, est-ce que le nombre minimal de policiers requis pour capturer les voleurs avec probabilité p , le p -cop number, est inférieur à un entier k ?

est EXPTIME-difficile lorsque \mathcal{G} et k font partie de l'entrée.

Démonstration. Comme le jeu de k policiers et voleur s'écrit dans le modèle de la définition 3.2.1 et est EXPTIME-complet, il suit donc que la complexité de \mathcal{G} ne peut être moindre. Autrement, une solution polynomiale pour \mathcal{G} pourrait être utilisée pour résoudre le jeu de k policiers et voleurs. \square

On ne fait aucune assertion quant à la l'appartenance de ce problème à la classe EXPTIME, mais il appert que celui-ci ne peut faire partie d'une classe plus facile que EXPTIME.

3.4 Un résultat de stationnarité

Dans les jeux de policiers et voleurs traditionnels dans lesquels sont définis une relation \preceq_n (tels que le jeu de policier et voleur classique [39] et le jeu à k policiers [12]), il est utile de démontrer des résultats sur la convergence de la relation \preceq_n . On démontre que la suite devient *stationnaire*, soit qu'il existe un rang $N \in \mathbb{N}$ tel que pour tout nombre $n \geq N$ et toute paire de sommets $(u, v) \in V^2$, $u \preceq_n v \implies u \preceq_{n+1} v$. On note \preceq pour la partie stationnaire de la suite, c'est-à-dire $\preceq := \preceq_N$. Ce résultat est primordial pour la résolution des jeux de policiers voleurs puisqu'il garantit que la relation \preceq peut être calculée en temps fini.

La question de la stationnarité de la récurrence w_n n'est pas triviale. Par exemple, le triangle K_3 est clairement policier-gagnant au sens classique, on a même $r \preceq_1 c$ pour toute position du voleur r et c du policier. De plus, $\preceq_1 = \preceq_k$ pour tout $k > 1$. Cependant, lorsqu'on ajoute une probabilité de capture sur les sommets, disons $\frac{1}{M}$ pour un certain $M > 0$, alors, pourvu que les joueurs ne commencent pas sur le même sommet, après n tours le policier aura gagné avec probabilité $1 - (1 - \frac{1}{M})^n \approx 1 - \frac{1}{e^n}$ lorsque M et n sont grands. Bref, après n'importe quel nombre fini de tours n , le policier aura gagné avec une probabilité strictement plus petite que 1, bien qu'il est clair qu'il peut gagner avec probabilité au moins p pour n'importe quel $p \in [0, 1)$. On énonce, dans la proposition suivante, une borne supérieure sur le nombre minimal n nécessaire pour déterminer la probabilité de capture des voleurs dans un jeu infini, $\lim_{n \rightarrow \infty} w_n(s)$.

Proposition 3.4.1. *Soit \mathcal{G} un jeu de policiers et voleurs. Alors, pour tout état $s \in S$ la récurrence $w_n(s)$ définie à l'équation (3.3) est telle que :*

1. *si $w_{|S|}(s) = 0$, alors pour tout $k > 0$, $w_{|S|+k}(s) = 0$;*
2. *si $w_{|S|+1}(s) = w_{|S|}(s) = p \in (0, 1)$, alors pour tout $k > 0$, $w_{|S|+k}(s) = w_{|S|}(s) = p$;*

3. si $w_{|S|+1}(s) > w_{|S|}(s)$, alors $\exists p \in (0, 1]$ tel que $\lim_{k \rightarrow \infty} w_{|S|+k}(s) = p$.

Démonstration. Comme les joueurs utilisent des stratégies optimales déterministes et sans mémoire, une fois que l'action optimale à jouer à partir d'un état est déterminée celle-ci est invariante. Ainsi, en supposant que ces actions ont toutes été calculées, on peut définir une chaîne de Markov \mathcal{M} décrivant le jeu comme suit. L'ensemble des états $S^{\mathcal{M}}$ de la chaîne est inclus dans l'union disjointe de F , $(S \setminus F) \times A_{\text{cop}}$ et $(S \setminus F) \times A_{\text{rob}}$. Chaque état de $S^{\mathcal{M}}$ est soit un état de jeu qui n'est pas final ainsi que l'action optimale jouée par le joueur se trouvant à cet état ou simplement un état final. On définit la matrice de transition M de la chaîne à l'aide des fonctions de transition T_{cop} et T_{rob} du jeu de policiers et voleurs comme suit :

$$M(\mathbf{s}, \mathbf{s}') = \begin{cases} T_{\text{cop}}(s, a_{\text{cop}}^*, s'), & \text{si } \mathbf{s} = (s, a_{\text{cop}}^*) \text{ et } \mathbf{s}' = (s', a_{\text{rob}}^*) \text{ ou } \mathbf{s}' \in F \\ T_{\text{rob}}(s, a_{\text{rob}}^*, s'), & \text{si } \mathbf{s} = (s, a_{\text{rob}}^*) \text{ et } \mathbf{s}' = (s', a_{\text{cop}}^*) \text{ ou } \mathbf{s}' \in F \\ 1, & \text{si } \mathbf{s} = \mathbf{s}' \in F \\ 0, & \text{sinon.} \end{cases}$$

Ici, a_{rob}^* et a_{cop}^* dénotent les actions optimales à choisir par les voleurs et les policiers déterminées par les stratégies optimales. La marche est alternée : d'un état et l'action optimale à y jouer pour un joueur, elle transite vers un autre état et l'action optimale à y jouer pour le joueur adverse. Les boucles autour des états finaux de F sont nécessaires pour que M définisse une chaîne de Markov.

On définit un parcours P comme un chemin, fini ou infini, dans \mathcal{M} avec répétitions de sommets. Soit $P_{(s, a^*)}^F$ l'ensemble des parcours finis de \mathcal{M} débutant en (s, a^*) et dont le seul élément de F est le dernier état du chemin. Comme les actions optimales déterminent la chaîne \mathcal{M} et que la récurrence w_n calcule la probabilité de capture des voleurs en au plus n tours, alors $w_n(s)$ cumule les probabilités de tous les chemins de longueur au plus $2n$ dans $P_{(s, a^*)}^F$. On note aussi que, puisque la chaîne est alternée, $|S^{\mathcal{M}}|$ pas dans la chaîne correspondent à $|S|$ tours de jeu. Alors, trois cas sont possibles.

Cas 1 : $P_{(s, a^*)}^F = \emptyset$. Les actions optimales des policiers et des voleurs ne permettent pas d'atteindre F en au plus $|S|$ pas dans \mathcal{G} et $w_n(s) = 0$ pour tout $n \geq 0$. Inversement, si $w_{|S|}(s) = 0$, alors $P_{(s, a^*)}^F$ ne contient aucun parcours de longueur au plus $2|S|$. Tout chemin P de $P_{(s, a^*)}^F$ de longueur plus de $2|S|$ contient un sous-chemin de longueur au plus $2|S| = |S^{\mathcal{M}}|$ puisque P doit contenir des répétitions. En effet, on peut enlever les cycles formés entre deux occurrences d'un même sommet et obtenir un chemin de longueur au plus $2|S|$. Ainsi, $P_{(s, a^*)}^F$ est non-vide si et seulement si il contient un chemin de longueur au plus $2|S|$. Comme $P_{(s, a^*)}^F$ est vide, on déduit que si $w_{|S|}(s) = 0$, alors $w_{|S|+k}(s) = 0$ pour tout $k \geq 1$.

Cas 2 : $P_{(s, a^*)}^F$ est fini. Alors, chaque chemin de $P_{(s, a^*)}^F$ ne contient aucun cycle. Autrement, $P_{(s, a^*)}^F$ contiendrait une suite de chemins $(P_i)_{i \in \mathbb{N}}$, où chaque P_i traverserait i fois le cycle. $P_{(s, a^*)}^F$

serait donc infini, ce qui est une contradiction. Donc, chaque chemin est de longueur au plus $2|S|$ et $w_{|S|}(s)$, qui cumule les probabilités de tous les chemins de longueur au plus $2|S|$ dans $P_{(s,a^*)}^F$, ne peut pas augmenter en $w_{|S|+1}(s)$. La probabilité de traverser un état final est donc stationnaire après au plus $|S|$ tours. On déduit que si $w_{|S|+1}(s) = w_{|S|}(s)$, alors $w_{|S|}(s) = w_{|S|+k}(s)$ pour tout $k > 0$.

Cas 3 : $P_{(s,a^*)}^F$ est infini. Certains parcours de $P_{(s,a^*)}^F$ doivent donc contenir des cycles dirigés. Ainsi, les policiers ont une infinité d'opportunités d'atteindre un état final et de capturer les voleurs tandis que ceux-ci peuvent s'échapper temporairement seulement puisqu'ils se retrouvent dans un état déjà rencontré. La probabilité de capture est donc strictement croissante, mais bornée par 1. Il existe donc $p \in (0, 1]$ tel que $w_{|S|+k}(s) \rightarrow p$ lorsque $k \rightarrow \infty$. Suite aux arguments précédents, on déduit que si la probabilité de capture n'est pas stationnaire, c'est-à-dire si $w_{|S|+1}(s) > w_{|S|}(s)$, celle-ci tend vers un nombre dans l'intervalle $(0, 1]$.

□

Dans la proposition précédente, dans le cas où $w_{2|S|}(s) > w_{|S|}(s)$ il serait impossible d'arrêter un calcul itératif à la valeur exacte de la probabilité de capture. En effet, on ne peut donner d'énoncé général que dans les deux premiers cas. Ceci est illustré dans l'exemple 3.4.2. On y voit qu'il existe une chaîne de Markov \mathcal{M} correspondant à un jeu de policiers et voleurs \mathcal{G} sur laquelle aucun calcul itératif ne peut en temps fini retourner la valeur exacte de la probabilité de capture des voleurs.

Exemple 3.4.2. Le graphe de la figure 3.2(a) illustre un exemple de jeu \mathcal{G} où les actions qui ne sont pas optimales ont été éliminées. Supposons qu'on y joue une variante du jeu du policier et du voleur classique avec probabilité de capture. La figure 3.2(a) montre une partie du graphe de jeu dans laquelle les joueurs sont arrivés à l'état s_0 . L'état *perdant* est un état dans lequel les deux joueurs sont sur des composantes connexes distinctes d'un graphe, auquel cas le voleur a gagné. Cet état est une abstraction qui agrège plusieurs états à partir desquels le policier ne peut pas gagner. À partir l'état s_0 , l'action c_0 , que l'on sait optimale, du policier le mène à l'état s_5 avec probabilité $T_{\text{cop}}(s_0, c_0, s_5) = \frac{1}{4}$, alors qu'avec probabilité $T_{\text{cop}}(s_0, c_0, s_1) = \frac{3}{4}$, il atteint l'état s_1 . Sur les deux représentations, des figures 3.2(a) et 3.2(b), les actions du policier sont notées c et celles du voleur r . C'est le seul exemple dans lequel cette notation est utilisée. À l'état s_1 , avec probabilité $T_{\text{rob}}(s_1, r_1, F) = \frac{1}{2}$, le voleur atteint un état de F , où il a perdu, avec l'action r_1 . Autrement, avec probabilité $T_{\text{rob}}(s_1, r_1, s_2) = \frac{1}{2}$, il atteint l'état s_2 . F est aussi une abstraction qui agrège tous les états dans lesquels le policier a gagné. Sur le cycle $(s_1, s_2, s_3, s_4, s_1)$ les joueurs sont à distance 1, mais le voleur ne peut se faire capturer qu'à son tour à l'état s_1 . La probabilité de capture du voleur tend donc vers $\frac{3}{4}$. Dans la figure 3.2(a), seules les actions optimales sont dessinées pour ne pas alourdir la représentation. Par exemple, aucune arête étiquetée c_4 , débutant en s_4 et terminant en s_2 n'est dessinée.

La chaîne de Markov de la figure 3.2(b) correspond au jeu illustré à la figure 3.2(a). On voit que les actions qui étaient sur des arêtes à la figure 3.2(a) sont maintenant incluses dans les états de la chaîne. Les cycles (s_1, \dots, s_4, s_1) , dessinés en bleu, sont les mêmes dans les deux représentations. L'état *perdant* est un état perdant pour le policiers, alors que F est un état gagnant. De l'état (s_0, c_0) , le policier joue l'action c_0 . Les sommets de la partition de droite représentent donc des sommets de la chaîne \mathcal{M} à partir desquels le voleur joue et inversement pour ceux de la partition de gauche. Pour ne pas alourdir le graphe, les états inutiles n'ont pas été dessinés à la figure 3.2(b) tels que les états (s_1, c_1) ou (s_2, r_2) puisque ceux-ci ne sont pas atteignables. Ici aussi, les états *perdant* et F sont des abstractions, le premier est un état qui agrège plusieurs états à partir desquels le policier ne peut pas gagner alors que le second agrège des états dans F . Les deux représentations des figures 3.2(a) et 3.2(b) se ressemblent beaucoup puisque nous avons omis les actions qui ne sont pas optimales ainsi que les actions d'un joueur qui ne peut atteindre l'état en question. Le chaîne \mathcal{M} ne garde que l'information pertinente au calcul de la récurrence w_n .

On voit sur la chaîne \mathcal{M} de la figure 3.2(b) que la probabilité d'atteindre l'état gagnant à partir de l'état (s_0, c_0) tend vers $\frac{3}{4}$ lorsque le nombre d'incrément de temps augmente. En effet, la chaîne contient deux états absorbants, deux états d'où on ne peut sortir, *perdant* et F , et la probabilité d'atteindre l'un d'eux tend vers 1 au fil du temps. Comme l'état (s_0, c_0) ne peut être retraversé, la probabilité d'atteindre chacun de ces états dépend de la transition initiale. Ainsi, la probabilité de victoire est $\frac{3}{4}$, mais $w_n(s_0) \rightarrow \frac{3}{4}$ quand $n \rightarrow \infty$.

À la lumière de la proposition 3.4.1, on voit qu'il est généralement possible de borner supérieurement le nombre minimal n suffisant dans le calcul de la récurrence $w_n(s)$ pour déduire si cette valeur est supérieure ou égale à un nombre $p \in [0, 1]$. Ainsi, on peut compléter la complexité algorithmique calculée à la proposition 3.3.1.

Corollaire 3.4.3. *En pire cas et en utilisant une approche de programmation dynamique, $O(|S|^3|A_{\text{cop}}||A_{\text{rob}}|)$ opérations sont suffisantes pour déterminer si $w_n(s)$ est nulle, stationnaire égale à un nombre $p \in (0, 1]$ ou non-stationnaire supérieure à p .*

Démonstration. La démonstration suit directement des propositions 3.4.1 et 3.3.1 en remplaçant n par $|S|$. On calcule une complexité temporelle de $O(|S|^3|A_{\text{cop}}||A_{\text{rob}}|)$. \square

Il a été utile dans la démonstration de la proposition 3.4.1 de remarquer que les jeux de policiers et voleurs peuvent être traités comme une chaîne de Markov une fois les stratégies optimales déterminées. L'intérêt de cette observation ne s'arrête pas là cependant. La théorie des chaînes de Markov est très vaste et contient de nombreux résultats profonds. Notamment, une des questions classiques de la théorie des chaînes de Markov concerne la détermination du temps de frappe espéré d'un état à partir d'un autre état. Cette question rejoint directement

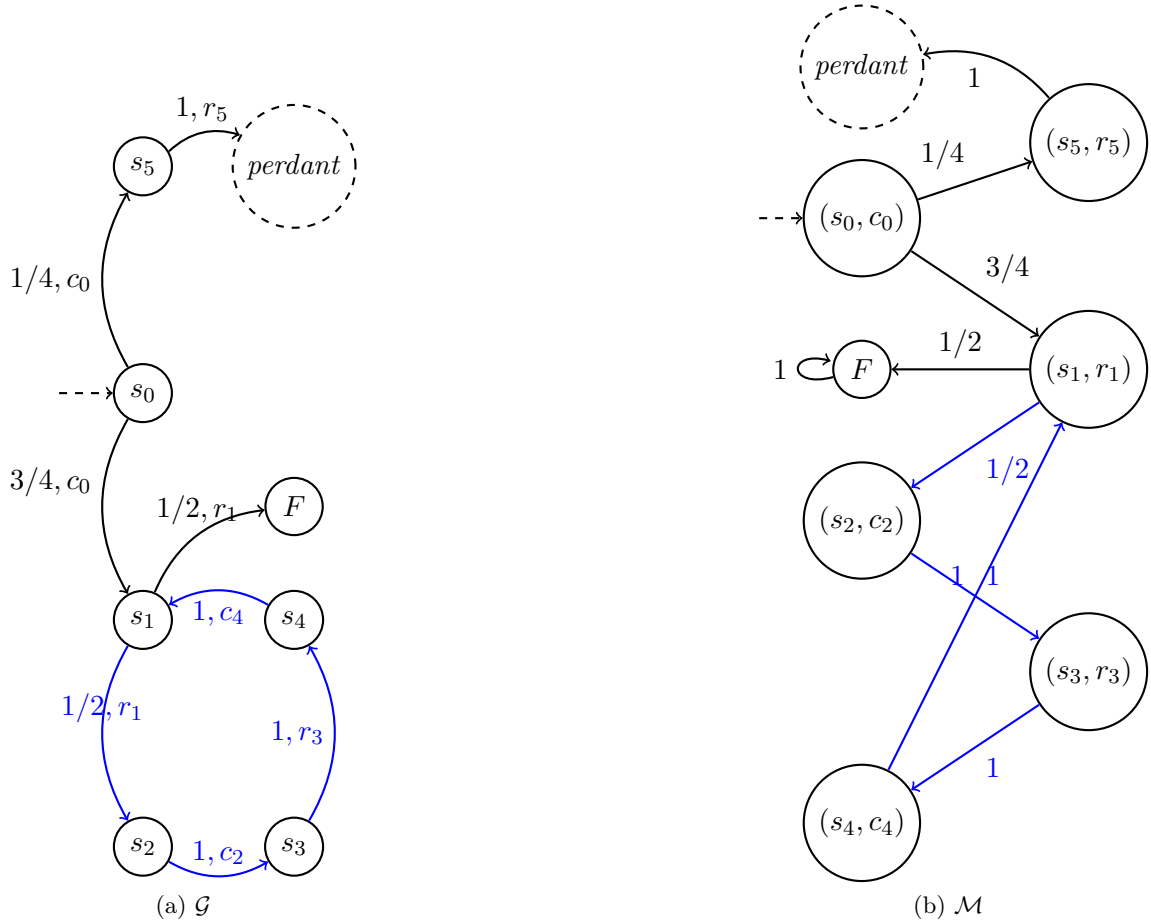


FIGURE 3.2 – Illustration du troisième cas de la proposition 3.4.1

une autre question, cette fois-ci dans les jeux de policiers et voleurs, celle du calcul du temps espéré de capture des voleurs³. On peut donc, sans résoudre directement cette question, donner quelques pistes de solution pour ce second problème.

Soit \mathcal{M} la même chaîne décrite à la démonstration de la proposition précédente 3.4.1 et M sa matrice de transition. Si $(X_n)_{n \geq 0}$ décrit le processus sur la chaîne \mathcal{M} débutant à l'état initial, alors on définit $T := \frac{1}{2} \min_{n \geq 0} (X_n \in F)$. T est donc le temps de capture des voleurs. L'espérance du temps de capture T en partant de $s \in S^{\mathcal{M}}$ est notée $\mathbb{E}[T | s]$. Un résultat assez élémentaire en chaînes de Markov est le suivant, que l'on présente sans démonstration. La démonstration de cette proposition peut être retrouvée dans le livre de Norris [38] sur les chaînes de Markov.

3. Dans les premiers articles décrivant le jeu de policier et voleur saoul [25, 28] les auteurs cherchaient à calculer des bornes supérieures sur l'espérance du temps de capture du voleur par le policier. On peut voir cette question comme un problème de la théorie du contrôle. Imaginons un agent pris dans un système évoluant de manière markovienne et sur lequel il peut exercer un certain contrôle, en moyenne combien de temps devra-t-il prendre pour forcer le système à entrer dans un certain état ?

Proposition 3.4.4 (Norris [38]). $\mathbb{E}[T \mid s]$ est la solution minimale non négative au système d'équations linéaires suivant, pour tout $s \in S^{\mathcal{M}}$:

$$\mathbb{E}[T \mid s] = \begin{cases} 0, & \text{si } s \in F; \\ 1 + \sum_{s' \in S^{\mathcal{M}}} M_{ss'} \mathbb{E}[T \mid s'], & \text{si } s \notin F. \end{cases}$$

$$= \begin{cases} 0, & \text{si } s \in F; \\ \frac{1}{1 - M_{ss}} \left(1 + \sum_{s' \in S^{\mathcal{M}} \setminus (F \cup \{s\})} M_{ss'} \mathbb{E}[T \mid s'] \right), & \text{si } M_{ss} \neq 1, s \notin F; \\ \infty, & \text{si } M_{ss} = 1, s \notin F. \end{cases}$$

Ainsi, une fois les stratégies optimales déterminées on peut simplement résoudre un système d'équations linéaires pour déterminer les temps de capture moyens à partir de n'importe quel état. Bien qu'elle nécessite une hypothèse apparemment forte, soit la connaissance complète de la chaîne \mathcal{M} , la proposition 3.4.4 peut être intéressante s'il est possible de calculer en temps réel chaque valeur de M dans le système linéaire. Autrement dit, si M n'a pas à être gardée complètement en mémoire, il est possible que la proposition précédente puisse être utilisée efficacement. D'autres résultats plus complexes existent dans la théorie des chaînes de Markov qui permettraient probablement de calculer $\mathbb{E}[T \mid s]$ plus efficacement. Certains requièrent cependant une analyse plus en détail de la chaîne \mathcal{M} . On cite ici simplement quelques articles [11, 13, 42] récents traitant de l'espérance du temps de frappe, correspondant ici à $\mathbb{E}[T \mid s]$ d'une chaîne de Markov quelconque.

Un autre aspect intéressant de la chaîne \mathcal{M} est la possibilité de réécrire la proposition 3.4.1 directement en terme de sa structure. En effet, la démonstration de la proposition montre que si l'on peut caractériser la structure de la chaîne \mathcal{M} , alors on peut aboutir aux mêmes conclusions de cette proposition sans calculer directement la récurrence w_n (sauf possiblement pour déterminer les actions optimales à jouer).

Corollaire 3.4.5. Soit $s \in S$ un état de jeu et \mathbf{s} son état correspondant dans \mathcal{M} . Supposons que \mathcal{M} est construite. Soit X l'ensemble des états de \mathcal{M} qui peuvent atteindre un état de F . Alors, trois cas sont possible.

1. $\mathbf{s} \notin X$, auquel cas $w_n(s) = 0$ pour tout $n \geq 0$.
2. $X \setminus F$ ne contient aucun cycle atteignable à partir de \mathbf{s} , auquel cas $(w_n(s))_{n \in \mathbb{N}}$ devient stationnaire.
3. $X \setminus F$ contient un cycle atteignable à partir de \mathbf{s} , auquel cas $(w_n(s))_{n \in \mathbb{N}}$ ne devient jamais stationnaire.

Démonstration. On reprend la définition de parcours $P_{(s,a^*)}^F$ de la démonstration de la proposition 3.4.1.

- Cas 1 : $\mathbf{s} \notin X$, alors par définition s ne peut pas atteindre d'état final. Ceci correspond au premier cas de la proposition 3.4.1 et $w_n(s) = 0$ pour tout $n \geq 0$.
- Cas 2 : $X \setminus F$ ne contient aucun cycle atteignable à partir de \mathbf{s} . Alors $P_{(s,a^*)}^F$ est fini selon le même argument que dans le second cas de la proposition 3.4.1. Ainsi, $(w_n(s))_{n \in \mathbb{N}}$ devient stationnaire.
- Cas 3 : $X \setminus F$ contient un cycle atteignable à partir de \mathbf{s} . Comme X ne contient que des états qui atteignent F , tous ces états ont un chemin positif vers F et l'argument du cycle qui fait augmenter la probabilité s'applique. Alors, $P_{(s,a^*)}^F$ est infini et $(w_n(s))_{n \in \mathbb{N}}$ ne devient jamais stationnaire.

□

Le corollaire 3.4.5 présente en quelque sorte une caractérisation structurelle des jeux de poursuite. Évidemment, ce résultat n'est pas aussi profond que les caractérisations classiques par démantèlement qui existent dans les jeux de policiers et voleurs [33]. Par contre, le corollaire affirme qu'une fois la chaîne de Markov \mathcal{M} décrite, alors il suffit d'analyser sa structure pour déduire les propriétés de la récurrence w_n de l'équation (3.3). Le corollaire 3.4.5 ne constitue pas une vraie caractérisation structurelle des jeux de policiers et voleurs puisqu'il n'est pas indépendant du calcul de la récurrence (3.3), mais c'est un premier jalon dans cette direction.

Remarque 3.4.6. *Une fois la chaîne \mathcal{M} construite, l'ensemble X peut être construit avec une méthode de parcours de graphe en profondeur à rebours en partant des états F . Autrement dit, soit A la matrice d'adjacence telle que :*

$$A(\mathbf{s}', \mathbf{s}) = \begin{cases} 1, & \text{si } M(\mathbf{s}, \mathbf{s}') > 0, \\ 0, & \text{sinon.} \end{cases}$$

Alors, pour chaque état $f \in F$, on peut faire une fouille de graphe en suivant la matrice A et inclure chaque état rencontré dans X . On note que A décrit un graphe dirigé dont les arêtes sont inversées par rapport au graphe sous-jacent à la chaîne \mathcal{M} . Une fois X construit, les cycles de $X \setminus F$ peuvent être découverts avec une fouille en profondeur.

On conclut cette section en notant qu'il reste à construire la chaîne de Markov \mathcal{M} . Sans aller plus loin, on note que celle-ci peut aisément être construite en pire cas en $O(|S|^4 |A_{\text{cop}}| |A_{\text{rob}}|)$ opérations, le temps suffisant pour calculer la récurrence $w_n(s)$ jusqu'à stationnarité sur chaque état $s \in S$.

3.5 Un modèle concret pour les jeux de policiers et voleurs

Dans cette section on définit un modèle de jeux de policiers et voleurs qui est plus concret et plus près des définitions usuelles dans la littérature. Ainsi, on précise que le jeu se déroule sur

un graphe, sans toutefois préciser la forme de celui-ci. Les actions des joueurs vont correspondre à des chemins, un peu comme dans le jeu du policier et du voleur rapide [33]. Le jeu présenté à la définition 3.2.1 est en effet abstrait puisque ses ensembles d'états et d'actions ne sont pas définis selon une structure précise. Notamment, l'analyse de la complexité algorithmique de l'équation w_n (3.3) ne dépend donc pas d'une structure précise, mais de ces ensembles abstraits. L'intérêt d'avoir un modèle défini sur un graphe est donc de pouvoir raffiner certains résultats en les rédigeant en fonction de la structure du graphe.

Dans le jeu présenté ci-dessous, les joueurs se déplacent le long de chemins puisqu'il semble, au regard de la littérature, que ces actions sont les plus générales possibles. En effet, les chemins permettent aux joueurs de se déplacer plus ou moins rapidement. Finalement, on accorde une zone de surveillance aux policiers leur permettant de capturer les voleurs autrement qu'en se déplaçant sur leurs sommets.

Définition 3.5.1. *Soit \mathcal{G} un jeu de policiers et voleurs tel que défini par la définition 3.2.1. Alors, un jeu \mathcal{H} de policiers et voleurs sur un graphe $G = (V, E)$ est un cas particulier de \mathcal{G} dans lequel k policiers et l voleurs s'affrontent sur G . \mathcal{H} est décrit par le tuple :*

$$\mathcal{H} := (S, I, F, A, T_{\text{cop}}, T_{\text{rob}}).$$

Les conditions suivantes doivent être respectées :

1. $S = S_{\text{cop}} \times S_{\text{rob}} \times S_o$ est l'ensemble fini des configurations de jeu.
2. $S_{\text{cop}} \subset V^k \times \mathcal{P}(E)^k$ est l'ensemble des configurations de jeu des policiers. V^k contient les positions des k policiers et $\mathcal{P}(E)^k$ leurs zones de surveillances.
3. $S_{\text{rob}} \subset V^l$ est l'ensemble des configurations de jeu des voleurs. V^l contient les positions possibles des l voleurs.
4. $A = A_{\text{cop}} \cup A_{\text{rob}}$ est l'ensemble fini des actions des policiers et des voleurs.
5. $A_{\text{cop}}(s) \subset \prod_{i=1}^k \mathcal{P}_{s_{\text{cop}}^i} \times \mathcal{P}(E)$ est l'ensemble des actions possibles des policiers. Chacun peut se déplacer le long d'un chemin et choisir une zone de surveillance.
6. $A_{\text{rob}}(s) \subset \prod_{i=1}^l \mathcal{P}_{s_{\text{rob}}^i}$ est l'ensemble des actions possibles des voleurs. Chacun peut se déplacer le long d'un chemin.
7. $T_{\text{cop}} : S \times A_{\text{cop}} \times S \rightarrow [0, 1]$ est une fonction satisfaisant $\sum_{s' \in S} T_{\text{cop}}(s, a, s') \in \{0, 1\}$ pour tout $s \in S$ et $a \in A_{\text{cop}}$. L'action a peut être jouée à partir de l'état s si et seulement si cette valeur est 1. Cette fonction est spécifique aux policiers.
8. $T_{\text{rob}}(s, a, s')$ est une fonction similaire à $T_{\text{cop}}(s, a, s')$, mais spécifique aux voleurs.

L'ensemble des remarques et détails qui s'ajoutent à la définition 3.2.1 sont encore applicables dans la définition précédente. Les policiers forment une équipe de k agents, alors que l voleurs se promènent sur le graphe. Une particularité ici est qu'on laisse les policiers avoir chacun une zone de surveillance constituant un ensemble d'arêtes. L'intérêt est de permettre à chaque

policier de capturer un ou plusieurs voleurs durant l'action de ce dernier. En effet, les voleurs se promènent le long de chemins. C'est un choix de modélisation naturel qui facilite le calcul de la probabilité de capture puisqu'on peut clairement voir quand un voleur est menacé de se faire capturer, cet événement ne se produisant que lorsqu'un voleur traverse la zone de surveillance d'un policier. Autrement, si les policiers surveillaient des sommets, la probabilité de capture des voleurs serait difficile à définir en fonction de leur action. La probabilité de transition des voleurs doit donc contenir la probabilité que ceux-ci survivent sur chacune des arêtes des chemins qu'ils ont choisi comme action pour se rendre à l'état s' puisque chaque fois qu'une telle arête est dans une zone de surveillance, un voleur risque de se faire prendre.

Évidemment, cette dernière définition est encore une fois modulaire, dans plusieurs jeux les policiers n'ont aucune zone de surveillance et on peut laisser tomber cette composante des états policiers.

Cette définition est illustrée sur l'exemple qui suit. Celui-ci décrit le jeu du policier et voleur rapide avec probabilité de capture qui est une variante du jeu de Marcoux [33]. Sans surprise, étant donné que ces deux jeux demandent aux voleurs de se déplacer le long de chemins, il est plus simple de décrire le jeu du voleur rapide en suivant la définition 3.5.1. Le jeu du voleur rapide a été présenté à l'exemple 3.2.6, mais sans probabilité de capture, tandis que le jeu du voleur évasif a été proposé à l'exemple 3.2.5. Le jeu suivant fusionne certains éléments de ces derniers.

Pour un chemin $\pi \in \mathcal{P}$ d'un graphe G , on écrit $\pi[k]$ pour dénoter son k^e sommet ainsi que $\pi[*]$ pour son dernier.

Exemple 3.5.2 (Le jeu du voleur rapide et évasif). Soit $G = (V, E)$ un graphe fini. Supposons que le policier surveille une zone $C \subseteq E$ et que chaque fois que le voleur emprunte une arête e il ressort vivant de sa traversée selon une probabilité $q_C(e)$. Dans l'exemple 3.2.5, on utilisait une probabilité de capture, alors qu'ici on utilise une probabilité de survie, qui est plus simple à utiliser dans le présent contexte. On suppose que $q_C(e) \geq 0$ pour toute arête e et $q_C(e) < 1$ si $e \in C$. Contrairement au jeu du voleur évasif 3.2.5, la probabilité de capture dépend de la zone surveillée par le policier ainsi que de l'action du voleur. Ici, seules la zone de surveillance du policier ainsi que les fonctions de transition des joueurs changent par rapport à l'exemple 3.2.6. Il existe donc un élément $j^* \notin V$ et l'ensemble d'états finaux est $F = \{(j^*, \emptyset, j^*)\}$. On écrit E_c pour l'ensemble des arêtes reliant un sommet c à son voisinage. Similairement, on écrit E_π pour dénoter l'ensemble des arêtes d'un chemin π . Par rapport au jeu de policier et voleur rapide de l'exemple 3.2.6, la fonction de transition du policier voit simplement sa composante $N[c]$ remplacée par E_c , autrement dit, le policier surveille maintenant les arêtes

contenues dans son voisinage. La fonction de transition du policier est :

$$T_{\text{cop}}((c, E_c, r), c') := \begin{cases} \delta_{(c', E_{c'}, r)}, & \text{si } (c = i_{\text{cop}}, c' \in V) \text{ ou si } (c \in V, c' \in N[c]) \\ 0, & \text{sinon.} \end{cases}$$

Celle du voleur, quant à elle, est :

$$T_{\text{rob}}((c, E_c, r), \pi) := \begin{cases} \delta_{\pi[*]}, & \text{si } E_\pi \cap E_c = \emptyset \\ D_{(r, \pi[*])}, & \text{si } E_\pi \cap E_c \neq \emptyset \end{cases}$$

où $D_{(r, \pi[*])}$ est une fonction satisfaisant :

$$D_{(r, \pi[*])}(x) := \begin{cases} \prod_{e \in E_\pi} q_{E_c}(e), & \text{si } x = (c, E_c, \pi[*]), \\ 1 - \prod_{e \in E_\pi} q_{E_c}(e), & \text{si } x = (j^*, \emptyset, j^*). \end{cases}$$

Puisque la zone de surveillance est déterminée par la position du policier, nous pouvons utiliser la notation simplifiée (c, r) pour un état (c, E_c, r) . Alors la récurrence de l'équation (3.3) s'écrit :

$$w_0(c, r) = 1 \iff (c, E_c, r) = (j^*, \emptyset, j^*);$$

$$w_n(c, r) = \begin{cases} 1, & \text{si } (c, E_c, r) = (j^*, \emptyset, j^*); \\ \max_{c' \in N[c]} \min_{\pi \in \mathcal{P}_r} T_{\text{rob}}((c', r), \pi, (c', \pi[*])) w_{n-1}(c', \pi[*]) \\ \quad + T_{\text{rob}}((c', r), \pi, (j^*, \emptyset, j^*)), & \text{autrement.} \end{cases}$$

Selon la proposition 3.4.1, la complexité algorithmique de la récurrence précédente requiert en pire cas $O(n|V|^2\Delta|\mathcal{P}|) = O(n\Delta|V|^4|\mathcal{P}|)$, où Δ est le degré maximal du graphe G . En effet, S correspond à l'ensemble des paires de sommets, le policier ne peut se déplacer que sur son voisinage à chaque tour et le voleur peut sélectionner n'importe quel chemin fini du graphe. Cependant, cette complexité peut maintenant être évaluée en fonction de la taille du graphe puisqu'on utilise un modèle défini sur une telle structure. On montre que le facteur $|\mathcal{P}|$ peut être réduit.

Proposition 3.5.3. *Calculer $w_n(i)$ pour le jeu précédent nécessite $O(n\Delta|E|(|E| + |V| \log |V|)) \subset O(n\Delta|V|^4)$ opérations et $O(n|V|^2)$ espace mémoire.*

Démonstration. On écrit \mathcal{P}_A^B pour l'ensemble des chemins débutant en A et terminant en B . Fixons la position (c, E_c) , du policier. L'action π du voleur peut être décomposée selon chacune des arêtes $e \in E_\pi$. Supposons que $q_{E_c}(e) = 1$ si $e \notin E_c$, pour simplifier la fonction de

transition du voleur. Alors, la récurrence précédente, lorsque l'état n'est pas final, peut être réécrite comme :

$$w_n(c, r) = \max_{c' \in N[c]} \min_{\substack{r' \in V \\ \pi \in \mathcal{P}_{r'}^{r'}}} \left(\prod_{[u,v] \in E_\pi} q_{E_c}([u,v]) w_{n-1}(c', v) + 1 - \prod_{[u,v] \in E_\pi} q_{E_c}([u,v]) \right).$$

Traverser l'ensemble des voisins de la position c du policier nécessite Δ opérations. Déterminer la destination r' du voleur nécessite de traverser l'ensemble des sommets de V .

Pour chaque destination r' , le voleur peut calculer le chemin le plus court entre r et r' en plaçant un poids $-\log q_{E_c}([u,v])$ sur chaque arête $[u,v] \in E$, correspondant à la probabilité de capture sur cette arête. Ceci se fait sans perte de généralité. En effet, le chemin π minimisant la fonction objectif $1 - \prod_{e \in E_\pi} q_{E_c}(e)$ est le même que celui maximisant la fonction objectif $\prod_{e \in E_\pi} q_{E_c}(e)$. Ce dernier est aussi le même que le chemin maximisant $\sum_{e \in E_\pi} \log q_{E_c}(e)$ puisque \log est une fonction monotone croissante, le maximum de la fonction $\sum_{e \in E_\pi} \log q_{E_c}(e)$ est atteint au même point que celui de la fonction $\prod_{e \in E_\pi} q_{E_c}(e)$. Comme $\log q_{E_c}(e) < 0$, puisque $q_{E_c}(e) \in [0, 1]$, on peut minimiser son inverse additif en préservant l'optimalité pour éviter les cycles négatifs.

Déterminer le chemin de longueur minimale entre r et l'ensemble des sommets du graphe nécessite en pire cas $O(|E| + |V| \log |V|)$ opérations, par exemple en utilisant l'algorithme de Fredman et Tarjan [18]. Ceci fait, la longueur du chemin de longueur minimale doit être au plus $|E| - 1$. Il reste à calculer $\prod_{[u,v] \in E_\pi} q_{E_c}([u,v]) w_{n-1}(c', v)$, ce qui nécessite au plus $|E| - 1$ calculs de $w_{n-1}(c', v)$. Si a_n dénote encore une fois le nombre d'opérations requis pour calculer w_n , alors on a :

$$\begin{aligned} a_0 &= 1 \\ a_n &\leq \Delta |E| O(|E| + |V| \log |V|) + a_{n-1} \\ a_n &\in O(n \Delta |E| (|E| + |V| \log |V|)), \end{aligned}$$

lorsque les valeurs de w_{n-1} sont sauvegardées en mémoire. □

Dans le cas présent, le policier était forcé de n'avoir comme zone de surveillance que son voisinage fermé. On pourrait toutefois imaginer que celui-ci puisse choisir des zones de surveillance plus complexes, par exemple en surveillant un sous-ensemble d'arêtes formant une clique dans son voisinage.

Un aspect important du jeu avec voleur(s) rapide(s) est sa capacité à modéliser des situations à information imparfaite dans lesquelles les policiers n'ont des informations sur la position des voleurs qu'à intervalle fixe. Le jeu avec témoins décrit ce jeu dans lequel à chaque k tours, pour k défini a priori, le policier observe la position du voleur. Autrement, le policier n'a aucune information sur les déplacements de son adversaire. Chalopin et coll. [10] présentent

une correspondance entre les versions complètement déterministes des jeux de policier et voleur rapide et du jeu avec témoins. Essentiellement, les auteurs démontrent que la classe des graphes policier-gagnants dans le jeu avec voleur rapide est la même que la classe des graphes policier-gagnants dans le jeu avec témoins. On peut donc s'interroger sur le cas stochastique pour voir si cette même correspondance tient toujours. Si tel est le cas, alors comme le jeu avec voleur rapide est résoluble en temps polynomial on peut directement prétendre résoudre un jeu à information imparfaite en temps polynomial. Ceci peut être très utile pour le développement de bornes telles que la borne COPWIN présentée au chapitre 4.

3.6 Exemples

Pour illustrer les modèles de jeu de policiers et voleurs des définitions 3.2.1 et 3.5.1, cette section présente quelques exemples dont certains jeux de policiers et voleurs classiques.

Exemple 3.6.1 (Le jeu de policier et voleur classique). Le jeu de Nowakowski et Winkler, et Quiliot, est présenté ici sous la forme du modèle 3.5.1. Soit $G = (V, E)$ un graphe fini, non dirigé, réflexif et connexe. L'ensemble d'états est $S := V^2$, on écrit d'ailleurs $s = (c, r)$ pour les positions du policier et du voleur. Les actions des deux joueurs sont simplement les voisins de leurs positions courantes, c'est-à-dire $N[c]$ pour le policier et $N[r]$ pour le voleur. Les états finaux sont ceux dans lesquels le policier et le voleur partagent le même sommet, $F = \{(c, r) \in S : c = r\}$. I est $\{(i_{\text{cop}}, i_{\text{rob}})\}$ et les actions à partir de I sont des choix de sommets, c'est-à-dire $A_{\text{cop}}(i_{\text{cop}}, i_{\text{rob}}) = A_{\text{rob}}(c, i_{\text{rob}}) = V$, $c \in V$. Finalement, les probabilités de transition sont triviales, l'état suivant est simplement le même sommet que l'action choisie.

Maintenant, pour montrer que la récurrence (3.3) est adéquatement définie, on démontre comment elle encode la relation \preceq_n de Nowakowski et Winkler caractérisée à la définition 2.1.1.

Le jeu est complètement déterministe, il est donc inutile d'écrire explicitement les fonctions de transition. On peut donc simplement écrire

$$\begin{aligned} w_0(c, r) &= 1 \iff c = r \\ w_n(c, r) &= \max_{c' \in N[c]} \min_{r' \in N[r]} w_{n-1}(c', r'). \end{aligned} \tag{3.5}$$

Il ne reste qu'à démontrer que la récurrence (3.5) simule exactement la relation \preceq_n . C'est ce qui est démontré dans la prochaine proposition.

Proposition 3.6.2. $w_n(c, r) = 1$ si et seulement si $\exists c' \in N[c] : r \preceq_n c'$.

Démonstration. On démontre le résultat par induction. On note que dans la récurrence w_n c'est au tour du policier à jouer, alors que dans la relation \preceq_n c'est au tour du voleur.

- $n = 0$: $w_0(c, r) = 1 \iff r = c \iff r \preceq_0 c$.
- $n = k + 1$: Supposons que le résultat tienne pour $n \leq k$ et démontrons le pour $n = k + 1$. $w_{k+1}(c, r) = 1$ si et seulement si il existe une action c' du policier telle que peu importe la réponse r' du voleur, $w_k(c', r') = 1$. Par l'hypothèse d'induction, $w_k(c', r') = 1 \iff \exists c'' \in N[c'] : r' \preceq_k c''$. Ainsi, le policier peut jouer l'action $c' \in N[c]$ et on a $r \preceq_{k+1} c'$. Inversement, si $\exists c' \in N[c] : r \preceq_{k+1} c'$ alors, par définition, pour toute action $r' \in N[r]$ du voleur il existe une réponse $c'' \in N[c']$ telle que $r' \preceq_k c''$. Par l'hypothèse d'induction, on a donc $w_k(c', r') = 1$. Dans ce cas, le policier peut jouer l'action $c' \in N[c]$ telle que, peu importe la réplique $r' \in N[r]$ du voleur, $w_k(c', r') = 1$. Par définition, on a donc $w_{k+1}(c, r) = 1$.

□

Exemple 3.6.3 (Le jeu de policier et voleur saoul). Il est clair que le jeu de policier et voleur saoul est une généralisation du jeu de policier et voleur classique, on présente donc ici la particularité de ce jeu par rapport au jeu classique. On reprend donc l'exemple précédent. Cependant, cette fois-ci la probabilité de transition du voleur est explicitée.

On peut modéliser le mouvement aléatoire du voleur à l'aide de sa fonction de transition T_{rob} . Pour n'importe quel état $(c, r) \in S$ et action $r' \in N[r]$ du voleur on écrit celle-ci comme :

$$T_{\text{rob}}((c, r), r', (c, r')) := \frac{1}{\deg(r) + 1}, \text{ pour tout } r' \in N[r].$$

Ainsi, l'action r' est insignifiante lorsque la marche aléatoire du voleur est uniforme et la fonction de transition décrit complètement la marche aléatoire du voleur.

On peut démontrer qu'il est toujours plus difficile de capturer un voleur jouant optimalement qu'un voleur jouant aléatoirement. Pour plus de généralité, on suppose que le voleur peut jouer selon n'importe quelle distribution dans $\text{Dist}_{N[r]}$ lorsqu'il se trouve sur le sommet r . Soit $\phi \subseteq (\text{Dist}_{N[r]})_{r \in V}$ un ensemble des distributions sur les sommets de V et ϕ_r la composante de ϕ qui est dans $\text{Dist}_{N[r]}$. Alors, on écrit $w_n^\phi(c, r)$ pour la récurrence dans laquelle $T_{\text{rob}}((c, r), r') = \phi_r$. Autrement dit, on écrit

$$w_n^\phi(c, r) = \max_{c' \in N[c]} \sum_{r' \in N[r]} \phi_r(r') w_{n-1}^\phi(c', r'),$$

si $c \neq r$ et $w_n^\phi(c, r) = 1$ si $c = r$, $\forall n \geq 0$. La récurrence du jeu classique (3.5) est notée $w_n(c, r)$.

Proposition 3.6.4. *Il est toujours plus facile de capturer un voleur jouant aléatoirement qu'un voleur optimal, autrement dit*

$$w_n^\phi(c, r) \geq w_n(c, r).$$

Démonstration. On écrit $\delta_{N[r]}$ l'ensemble des distributions Dirac définies sur $N[r]$. Le voleur serait plus difficile à capturer s'il minimisait sa probabilité de capture seulement sur des

distributions Dirac puisque sa stratégie optimale est déterministe (voir la section 2.2 ou le théorème A.1.4). Alors, on calcule :

$$\begin{aligned}
w_n^\phi(c, r) &:= \max_{c' \in N[c]} \sum_{r' \in N[r]} \phi_r(r') w_{n-1}^\phi(c', r') \\
&\geq \max_{c' \in N[c]} \min_{\psi \subseteq (\text{Dist}_{N[r]})_{r \in V}} \sum_{r' \in N[r]} \psi_r(r') w_{n-1}^\psi(c', r') \\
&= \max_{c' \in N[c]} \min_{\psi \subseteq (\delta_{N[r]})_{r \in V}} \sum_{r' \in N[r]} \psi_r(r') w_{n-1}^\psi(c', r') \\
&= \max_{c' \in N[c]} \min_{r' \in N[r]} w_{n-1}(c', r').
\end{aligned}$$

La première ligne est la définition de la récurrence w_n lorsque le voleur est forcé de jouer selon une distribution ϕ . S'il pouvait choisir cette distribution, il pourrait trouver une distribution $\psi \subseteq (\text{Dist}_{N[r]})_{r \in V}$ lui assurant une probabilité de capture plus faible aux prochains tours, justifiant la seconde ligne. Ensuite, comme sa stratégie optimale est déterministe, elle correspond à un ensemble de distributions Dirac et le voleur ne perd rien à choisir $\psi \subseteq (\delta_{N[r]})_{r \in V}$. La dernière ligne est une réécriture de l'avant-dernière, car dans ce cas ψ_r est concentrée sur un sommet $r' \in N[r]$. \square

Exemple 3.6.5 (Le jeu sur graphe variable). En théorie des graphes, on peut définir différents processus aléatoires pour générer de manière stochastique des graphes se déformant à chaque incrément de temps. On peut donc obtenir une suite de graphes G_0, G_1, \dots représentant l'évolution d'un graphe en fonction du temps. Ces graphes sont aussi nommés *graphes dynamiques*. Ceux-ci peuvent modéliser, par exemple, la destruction d'un pont ou d'une route rendant impossible sa traversée par les policiers et les voleurs.

Pour prendre en compte la nature variable de la structure sous-jacente au jeu dans la définition 3.2.1, on peut simplement utiliser la composante S_0 comme compteur de tours. Comme le jeu se déroule sur des graphes, on illustre ceci sur la définition 3.5.1. Soit $G_t = (V_t, E_t)$ le graphe généré au temps t , alors on pose :

$$S = \left(\bigcup_{t=1}^{\infty} V_t^k \times \mathcal{P}(E_t)^k \right) \times \bigcup_{t=1}^{\infty} V_t^l \times \mathbb{N}.$$

Ainsi, à chaque incrément de temps t un nouveau graphe G_t est utilisé selon un certain processus aléatoire et l'ensemble d'états est renouvelé. Les ensembles d'actions sont réécrits en conséquence. On définit $\mathcal{P}_u^{G_t}$ l'ensemble des chemins de longueur finie du graphe G_t débutant à un sommet $u \in V_t$. Les ensembles d'actions sont :

$$\begin{aligned}
A_{\text{cop}}(c, C, r, t) &\subset \prod_{i=1}^k \mathcal{P}_{c_i}^{G_t} \times \mathcal{P}(E_t); \\
A_{\text{rob}}(c, C, r, t) &\subset \prod_{i=1}^l \mathcal{P}_{r_i}^{G_t}.
\end{aligned}$$

Comme cet exemple est assez général, après tout, seule la structure sous-jacente a été modifiée pour permettre au jeu de se dérouler sur un graphe dynamique, on laisse les fonctions de transitions indéfinies. On note toutefois qu'il faut faire attention avec ce type de jeu pour bien traiter le cas d'une action qui n'est pas définie dans le graphe généré au tour suivant, c'est-à-dire certains sommets d'une action a n'existeraient plus dans le nouveau graphe. Dans ce cas, on peut simplement piger au hasard une nouvelle action a' dont la distance avec a serait minimale. Ainsi, la nouvelle action ne dévierait pas trop du choix initial du joueur.

Exemple 3.6.6 (Stratégies à coût minimal). Dans certaines instances, on peut vouloir ajouter des coûts à chaque action que les joueurs peuvent prendre. Par exemple, certains policiers pourraient être plus efficaces que d'autres, mais seraient payés plus cher.

Pour modéliser ce type de scénario, on peut ajouter une fonction $c : A(s) \rightarrow \mathbb{R}$ au modèle 3.2.1 qui enverrait chaque action vers un coût. On ajouterait ensuite au modèle la seconde fonction objectif visant à minimiser le coût parmi l'ensemble des stratégies possibles. On obtiendrait un problème d'optimisation multi objectif visant à maximiser la probabilité de capturer les voleurs tout en minimisant les coûts. Ce problème pourrait même être décrit comme un problème d'optimisation biniveau : on pourrait chercher la stratégie des policiers la moins coûteuse sous la contrainte que celle-ci soit optimale. Une introduction à la théorie de l'optimisation biniveau peut être trouvée dans l'article de Colson et coll. [14]

Exemple 3.6.7 (Le jeu de Bonato et MacGillivray). Il a déjà été mentionné qu'un modèle général de jeux de policiers et voleurs a précédemment été présenté par Bonato et MacGillivray [6]. Ce modèle est cependant complètement déterministe et devrait donc pouvoir être aisément inclus dans notre modèle 3.2.1. Cet exemple fait donc la correspondance entre nos deux jeux.

Le modèle de Bonato et MacGillivray est présenté à la définition suivante.

Définition 3.6.8 (Le modèle de Bonato et MacGillivray). *Un processus à temps discret \mathcal{G} est un jeu de policiers et voleurs généralisé s'il satisfait aux règles suivantes :*

1. *Deux joueurs, poursuivant et évadé s'affrontent.*
2. *Il y a information parfaite.*
3. *Il y a un ensemble \mathcal{P}_P de positions admissibles pour le poursuivant et un ensemble \mathcal{P}_E pour l'évadé. L'ensemble des positions de jeu est le sous-ensemble $\mathcal{P} \subseteq \mathcal{P}_P \times \mathcal{P}_E$ de positions qui peuvent être atteintes selon les règles du jeu. L'ensemble des états de jeu est le sous-ensemble $S \subseteq \mathcal{P} \times \{P, E\}$ tel que $((p_P, q_E), X) \in S$ si, lorsque X est le prochain à jouer, la position (p_P, q_E) peut être atteinte en suivant les règles du jeu.*
4. *Pour chaque état de jeu et chaque joueur, il existe un ensemble non vide de mouvements alloués. Chaque mouvement laisse la position de l'autre joueur inchangée. On écrit $\mathcal{A}_P(p_P, q_E)$ pour dénoter l'ensemble de mouvements alloués au poursuivant lorsque l'état de jeu est $((p_P, q_E), P)$ et $\mathcal{A}_E(p_P, q_E)$ pour dénoter l'ensemble de mouvements alloués à l'évadé lorsque l'état de jeu est $((p_P, q_E), E)$.*

5. Les règles du jeu précisent comment celui-ci débute. Ainsi, il existe un ensemble $\mathcal{I} \subseteq \mathcal{P}_P \times \mathcal{P}_E$ de positions de départ admissibles. On définit $\mathcal{I}_P = \{p_P : \exists q_E \in \mathcal{P}_E, (p_P, q_E) \in \mathcal{I}\}$ et, pour $p_P \in \mathcal{P}_P$, on définit l'ensemble $\mathcal{I}_E(p_P) = \{q_E \in \mathcal{P}_E : (p_P, q_E) \in \mathcal{I}\}$. Le jeu \mathcal{G} commence avec le poursuivant qui choisit une position $p_P \in \mathcal{I}_P$ et ensuite l'évadé qui choisit une position $q_E \in \mathcal{I}_E(p_P)$.
6. Lorsque chaque joueur a choisi sa position initiale, le jeu se déroule alternativement avec le poursuivant jouant en premier. Chaque joueur, à son tour, doit choisir un mouvement admissible selon l'état de jeu courant.
7. Les règles du jeu précisent quand le poursuivant a capturé l'évadé. Autrement dit, il y a un sous-ensemble \mathcal{F} de positions finales. Le poursuivant gagne \mathcal{G} si, à n'importe quel moment, la position courante appartient à \mathcal{F} . L'évadé gagne si sa position n'appartient jamais à \mathcal{F} .

Seuls les jeux de policiers et voleurs dans lesquels l'ensemble \mathcal{P} de positions est fini sont considérés. Les jeux considérés sont joués sur une suite finie de tours qui sont indicés par des nombres naturels incluant 0.

Maintenant que le modèle 3.6.8 de Bonato et MacGillivray est présenté, on présente aussi comment ceux-ci ont défini une extension de la relation \preceq_n de Nowakowski et Winkler [39] pour résoudre l'ensemble des jeux caractérisés par leur modèle.

Définition 3.6.9 (La relation \preceq_n de Bonato et MacGillivray). *Soit \mathcal{G} un jeu de policiers et voleurs défini selon la définition 3.6.8. On pose*

1. $q_E \preceq_0 p_P$ si et seulement si $(p_P, q_E) \in \mathcal{F}$.
2. Supposons que $\preceq_0, \preceq_1, \dots, \preceq_{i-1}$ ont été définis pour un certains $i \geq 1$. Posons $q_E \preceq_i p_P$ si $(p_P, q_E) \in \mathcal{F}$ ou si $((p_P, q_E), E) \in \mathcal{S}$ et pour tout $x_E \in \mathcal{A}_E(p_P, q_E)$ soit $(p_P, x_E) \in \mathcal{F}$ ou il existe $w_P \in \mathcal{A}_P(p_P, x_E)$ tel que $x_E \preceq_j w_P$ pour un certain $j < i$.

Par définition \preceq_i contient \preceq_{i-1} pour chaque $i \geq 1$. Comme $\mathcal{P}_E \times \mathcal{P}_P$ est fini, il existe t tel que $\preceq_t = \preceq_k$ pour tout $k \geq t$. On pose $\preceq = \preceq_t$.

Bonato et MacGillivray utilisent ensuite la relation 3.6.9 pour démontrer une condition nécessaire et suffisante pour l'existence d'une stratégie gagnante pour le poursuivant qui est très similaire au théorème correspondant de Nowakowski et Winkler [39]. Ce théorème est présenté ici à des fins de complétude. La démonstration, relativement simple, peut être trouvée dans l'article de Bonato et MacGillivray [6]. On note que dans des jeux déterministes comme celui défini en 3.6.8, il est possible d'utiliser le terme de *stratégie gagnante*. Celle-ci est une stratégie déterministe garantissant au joueur de gagner contre n'importe quelle stratégie employée par son adversaire. Autrement dit, c'est une stratégie déterministe garantissant une valeur de 1 au joueur l'employant (en notation des jeux stochastiques). Bonato et MacGillivray définissent

implicitement les stratégies gagnantes comme étant sans mémoire. L'existence de stratégies gagnantes sans mémoire dans les jeux déterministes est garantie par plusieurs résultats [56].

Théorème 3.6.10 (Le théorème *policiers-gagnants* de Bonato et MacGillivray). *Le poursuivant a une stratégie gagnante dans un jeu de policiers et voleurs définis selon la définition 3.6.8 si et seulement s'il existe $p_P \in \mathcal{I}_P$ telle que pour tout $q_E \in \mathcal{I}_E(p_P)$, soit $(p_P, q_E) \in \mathcal{F}$ ou il existe $w_P \in \mathcal{A}_P(p_P, q_E)$ telle que $q_E \preceq w_P$.*

Clairement, ces deux jeux, celui de Bonato et MacGillivray (définition 3.6.8) et le nôtre (définition 3.2.1), décrivent des jeux de poursuite à information parfaite sur des structures discrètes. Les deux se déroulent aussi à tour de rôle. Bien que la notation soit différente dans les deux cas, on peut voir que le modèle de Bonato et MacGillivray peut être encodé dans le jeu décrit à la définition 3.2.1. La seule différence formelle entre nos deux formalismes concerne les états initiaux. En effet, Bonato et MacGillivray permettent l'utilisation d'ensemble d'états initiaux généraux I , alors que nous nous sommes restreints au cas où I ne contient qu'un seul élément. Ceci ne pose pas de problème puisqu'il suffit de jouer un tour de plus dans notre jeu et, possiblement, de modifier les premiers états atteignables pour avoir une correspondance exacte. Pour simplifier, dans ce qui suit on suppose que les ensembles d'états initiaux dans nos deux modèles sont équivalents. On conclut donc que l'équation (3.3) doit pouvoir encoder la relation de préordre \preceq_n de Bonato et MacGillivray. En effet, outre le caractère purement déterministe de la relation \preceq_n de la définition 3.6.9, on remarque que la structure de celle-ci est très similaire à celle de l'équation w_n de la définition 3.2.13. Aussi, les deux relations, \preceq_n et w_n , sont binaires et récursives. Ces deux relations sont aussi définies à l'aide de trois cas distincts : un premier cas lorsque $n = 0$ et dans lequel les joueurs n'ont aucune action à jouer ; un deuxième lorsque $n > 0$, mais que l'état courant est final ; finalement, un dernier cas, plus général, dans lequel les joueurs doivent jouer une action qui soit optimale en vue des tours subséquents. Bref, on démontre formellement que ces deux équations sont reliées dans les prochaines lignes.

On note en premier lieu que l'on peut simplifier l'équation (3.3) de la récurrence w_n lorsque l'on suit le caractère déterministe du jeu de Bonato et MacGillivray. La composante s_o n'étant pas utilisée dans ce qui suit, on écrira simplement $(c, r) \in S$. Comme $T_{\text{cop}}((c, r), a_{\text{cop}}) \in \delta_{S_{\text{cop}} \times \{r\}}$ pour tout $(c, r) \in S$, $a_{\text{cop}} \in A_{\text{cop}}(c, r)$ dans les deux modèles, on écrit $c' \in S_{\text{cop}}$ pour l'unique élément tel que $T_{\text{cop}}((c, r), a_{\text{cop}}, (c', r)) > 0$. De même, $r' \in S_{\text{rob}}$ est l'unique élément tel que $T_{\text{rob}}((c', r), a_{\text{rob}}, (c', r')) > 0$ puisque $T_{\text{rob}}((c', r), a_{\text{rob}}) \in \delta_{\{c'\} \times S_{\text{rob}}}$ pour tout $(c', r) \in S$ et $a_{\text{rob}} \in A_{\text{rob}}(c', r)$. La récurrence w_n est donc :

$$w_0(c, r) := 1 \iff (c, r) \in F;$$

$$w_n(c, r) = \begin{cases} 1, & \text{si } (c, r) \in F; \\ \max_{a_{\text{cop}} \in A_{\text{cop}}(c, r)} \min_{a_{\text{rob}} \in A_{\text{rob}}(c', r)} w_{n-1}(c', r'), & \text{autrement.} \end{cases}$$

Le théorème qui suit fait donc la correspondance exacte entre les deux formalismes que sont celui de Bonato et MacGillivray ainsi que celui qui est développé dans ce mémoire. Pour clarifier l'exposition, la relation \preceq_n est réécrite dans notre formalisme, celui de la définition 3.2.1. Il est apparent toutefois que la définition 3.2.1, notre modèle, englobe la définition 3.6.8 et cette réécriture se fait donc sans perte de généralité.

Théorème 3.6.11. *Soit la relation \preceq_n définie à la définition 3.6.9 et w_n la récurrence de l'équation (3.3). Supposons aussi que \mathcal{G} est un jeu de policiers et voleurs selon la définition 3.2.1, mais respectant les spécifications du modèle 3.6.8. Alors, on a*

$$w_n(c, r) = 1 \iff \exists a_{\text{cop}} \in A_{\text{cop}}(c, r) : r \preceq_n c', \forall (c, r) \in S. \quad (3.6)$$

Démonstration. On remarque que la relation \preceq_n compare des positions, position du poursuivant (policiers) et position de l'évadé (voleurs). Ces positions sont toutes deux encodées dans les états de jeu S de notre modèle. De plus, les ensembles d'actions \mathcal{A} définis dans le modèle 3.6.8 sont en fait des restrictions des ensembles d'actions de notre modèle 3.2.1. En effet, les actions dans le jeu de Bonato et MacGillivray correspondent directement à des positions de joueur, alors qu'on permet dans le modèle 3.2.1 aux ensembles d'actions d'être disjoints des ensembles d'états. Ainsi, il est possible de définir \mathcal{G} pour respecter les hypothèses du modèle 3.6.8 et donc la relation (3.6) est bien définie. Une différence subtile qu'il est important de noter entre l'équation (3.3) et la définition 3.6.9 concerne l'alternance des tours : les policiers sont les suivants à jouer dans la relation w_n alors que l'évadé (voleurs) joue le prochain dans la relation \preceq_n . Ceci ne change rien au fait que les policiers (poursuivant) jouent en premier dans les deux jeux.

Maintenant, on démontre par induction que l'équivalence tient.

- $n = 0$: $w_0(c, r) = 1 \iff (c, r) \in F \iff r \preceq_0 c$.
- $n = k+1$: Supposons que le résultat tienne pour $n \leq k$ et démontrons le pour $n = k+1$. $w_{k+1}(c, r) = 1$ si et seulement si $(c, r) \in F$, auquel cas $r \preceq_{k+1} c$ par définition, ou il existe une action $a_{\text{cop}} \in A_{\text{cop}}(c, r)$ des policiers telle que peu importe la réplique $a_{\text{rob}} \in A_{\text{rob}}(c', r)$ des voleurs, on a $w_k(c', r') = 1$. Par l'hypothèse d'induction, on a $w_k(c', r') = 1 \iff \exists a'_{\text{cop}} \in A_{\text{cop}}(c', r') : r' \preceq_k c''$. Ainsi, si les policiers jouent l'action a_{cop} , ils se placent dans une position dans laquelle $r \preceq_{k+1} c'$. Inversement, supposons qu'il existe une action $a_{\text{cop}} \in A_{\text{cop}}(c, r)$ telle que $r \preceq_{k+1} c'$. Alors, par définition, pour toute réplique $a_{\text{rob}} \in A_{\text{rob}}(c', r)$ des voleurs il existe une action $a'_{\text{cop}} \in A_{\text{cop}}(c', r')$ des policiers telle que $r' \preceq_k c''$. Dans ce cas, par l'hypothèse d'induction, on a $w_k(c', r') = 1$. Les policiers jouent l'action $a_{\text{cop}} \in A_{\text{cop}}(c, r)$, auquel cas on a $w_{k+1}(c, r) = 1$.

□

Exemple 3.6.12 (Le jeu des zombies). Les zombies ont acquis une notoriété fascinante dans les médias ces dernières années et sont souvent dépeints comme des êtres sans cervelle qui ne

cherchent qu'à se nourrir. On peut imaginer que ceux-ci jouent le rôle des policiers dans un jeu de policiers et voleurs alors que les survivants joueraient celui des voleurs. Récemment, Bonato et coll. [7] donnaient un exemple d'un jeu de zombies en modélisant ceux-ci comme des policiers jouant selon des stratégies sous-optimales simples contre un seul survivant. Plus précisément, chaque zombie est contraint de se déplacer sur un plus court chemin le séparant du survivant, si plus d'un tel chemin existe le zombie en pige un au hasard avec probabilité uniforme. Les auteurs cherchaient le nombre minimal de zombies requis pour assurer la capture du survivant lorsque le jeu est complètement déterministe.

On peut étendre facilement ce jeu en supposant que le survivant soit armé et puisse se défendre. Le survivant pourrait donc soit choisir de se déplacer, soit tuer un zombie. On pourrait imposer une zone de surveillance au survivant et restreindre le survivant à n'attaquer que les zombies traversant cette zone. On peut aussi imaginer qu'au lieu de choisir comme stratégie sous-optimale seulement un plus court chemin, les zombies à leur tour choisirait n'importe quelle action admissible sous-optimale a_{cop} avec probabilité p et une action optimale a_{cop}^* avec probabilité $1 - p$. Les survivants (voleurs) cherchent ici à atteindre un état de $S \setminus F$ duquel aucun état de F n'est atteignable. Par exemple, dans un tel état il se pourrait que tous les zombies soient morts (sous une définition adéquate de *mort* pour un zombie) ou immobilisés, auquel cas il leur serait clairement impossible d'atteindre un état final de F .

3.7 Conclusion et travaux futurs

En conclusion, ce chapitre a présenté un modèle relativement simple, mais très général pour décrire des jeux de policiers et voleurs. Celui-ci est présenté aussi avec une méthode de résolution sous la forme de la récurrence w_n . Ainsi, on prétend qu'une très grande variété de jeux de policiers et voleurs peuvent être résolus avec les objets développés dans ce chapitre, dont tous les jeux déjà connus dans la littérature.

Il reste, évidemment, du travail à faire non seulement pour avoir une théorie complète des jeux de policiers et voleurs, mais aussi pour avoir une théorie intéressante en pratique pour les chercheurs de différents domaines (mathématiques, informatique, robotique, recherche opérationnelle, etc.).

Notamment, une question qui reste encore ouverte porte sur la complexité de calcul de la récurrence. La complexité présentée à la section 3.3 laisse croire que certains jeux pourraient nécessiter un nombre exponentiel d'opérations. D'un autre côté, le modèle décrit par la définition 3.2.1 comprend le modèle présenté par Bonato et MacGillivray [6] qui est décrit comme un jeu combinatoire. Étend donné sa généralité, lui-même semble pouvoir décrire une grande quantité de jeux combinatoires, ce qui est un peu risqué puisqu'il est connu dans la littérature que plusieurs jeux combinatoires sont EXPTIME-complet à résoudre. C'est le cas, par exemple, du jeu de dames [47]. La classe de complexité EXPTIME-complet ne contient aucun

problème résoluble en temps polynomial, les problèmes s’y trouvant sont donc très difficiles à résoudre. Alors, une question intéressante est la suivante.

Comme certains jeux combinatoires font partie de la classe de complexité EXPTIME-complet, que le modèle de Bonato et MacGillivray semble incorporer plusieurs jeux combinatoires et que celui-ci est inclus dans le modèle 3.2.1, est-il vrai qu’en général résoudre un jeu décrit par le modèle 3.2.1 est EXPTIME-complet ?

On peut aussi se demander ce qu’il advient du paramètre qu’on appelle le *cop number* $c(G)$ dans les jeux de policiers et voleurs déterministes. Quoique sa définition ait été étendue aux jeux contenant des aspects stochastiques sous la forme du p -cop number, on ne peut en dire autant des résultats qui l’entourent. Ainsi, plusieurs années de recherche ont mené à la découverte de plusieurs résultats profonds décrivant différents aspects du paramètre $c(G)$ tel que son comportement sur des graphes aléatoires. D’ailleurs, la conjecture de Meyniel [8], selon laquelle $c(G) \in O(\sqrt{|V(G)|})$, n’est pas encore résolue. L’extension de cette conjecture au modèle 3.2.1 n’est pas triviale. On peut se demander, en effet, quelle fonction $f : \mathcal{G} \rightarrow \mathbb{N}$ décrirait le mieux le comportement asymptotique du paramètre $c_p(\mathcal{G})$, autrement dit telle que $c_p(\mathcal{G}) \in O(f(\mathcal{G}))$, où $c_p(\mathcal{G})$ dénote le p -cop number du jeu \mathcal{G} .

3.7.1 Jeux à information imparfaite

Enfin, une dernière avenue de recherche pouvant être envisagée après la lecture de ce chapitre et qui est probablement plus intéressante pour les chercheurs en robotique et en recherche opérationnelle concerne l’extension du modèle 3.2.1 aux jeux à information imparfaite. L’information imparfaite décrit le manque d’information de la part d’un ou des joueurs. Les jeux de policiers et voleurs à information imparfaite contiennent donc notamment les jeux avec voleurs invisibles qui peuvent modéliser exactement les problèmes de fouilles de graphes. En effet, tel que remarqué dans le chapitre 4 traitant de l’OSP, le problème de retrouver un objet perdu, mobile ou immobile, peut être décrit comme un jeu avec voleur invisible sous certaines conditions. La théorie des jeux semble permettre une transition relativement simple, en terme de modélisation, des jeux à information parfaite aux jeux à information imparfaite en forçant les joueurs à prendre leurs décisions non pas à partir d’états connus, mais d’ensemble d’états possibles.

En effet, on pourrait définir un ensemble $B \in \mathcal{P}(S)$ des ensembles d’états possible (B pour *belief state*, des états dont on croit possible). Cet ensemble contiendrait un certain nombre, généralement plus qu’un, d’états incluant l’état de jeu courant et décrirait l’ensemble des états de jeu dans lequel le jeu pourrait se trouver à un tour particulier. Les policiers n’auraient accès qu’à l’information comme quoi l’état de jeu courant est choisi parmi B . En parlant de jeux à information imparfaite, on suppose généralement que seuls les policiers ont de la difficulté à observer les voleurs, mais que ceux-ci connaissent la position des policiers. En s’inspirant de la théorie des jeux, on remarque qu’on peut donc décrire un jeu de policiers et voleurs à

information imparfaite en redéfinissant $A_{\text{cop}}(s)$ en $A_{\text{cop}}(b)$ pour un ensemble $b \in B$. Alors, les policiers ne prendraient pas de décision sur leur prochaine action en connaissant l'état courant, mais seulement l'ensemble contenant celui-ci. Ceci les empêcherait de déduire quel est l'état de jeu véritable, et donc de déduire la position des voleurs.

Bien qu'il resterait à formaliser le tout pour s'assurer que le modèle obtenu en étendant le modèle 3.2.1 soit cohérent et bien défini, il ne semble pas y avoir de difficulté majeure à cette étape-ci. L'obstacle majeur survient cependant lorsque vient le temps de résoudre le modèle obtenu. Déjà, on peut oublier l'idée d'obtenir un jeu résoluble en temps polynomial. Tel qu'il sera présenté dans le chapitre 4, un jeu aussi simple que le jeu du policier et voleur classique devient NP-complet lorsque le policier n'a aucune information sur la position de ce dernier. Cette généralisation, tel qu'il sera démontré au prochain chapitre, est plutôt intéressante pour le lien qu'elle crée avec le modèle 3.2.1 qui permettrait d'approximer des stratégies optimales dans les jeux avec information imparfaite à l'aide de méthodes de *branch and bound*.

3.7.2 En conclusion

À la lecture de plusieurs articles sur les jeux de policiers et voleurs, il peut sembler que ce chapitre se distance un peu de la ligne directrice typique établie dans ces jeux. Notamment, les chercheurs de ce domaine s'intéressent typiquement au nombre minimal de policiers nécessaire pour capturer le voleur, le *cop number*, nombre qui est redéfini à chaque variante conçue. Les solutions analytiques sont généralement privilégiées : rares sont les méthodes permettant d'estimer ce *cop number*, on y préfère des bornes asymptotiques. On peut présenter des caractéristiques de graphe permettant de reconnaître qu'un graphe est policier-gagnant ou non. On fait souvent l'hypothèse que les joueurs jouent de manière optimale et qu'aucun événement aléatoire ne peut se produire dans un jeu, autrement dit on suppose généralement que les jeux de policiers et voleurs sont des *jeux combinatoires*. De ces points de vue, on se distance en effet d'une partie de la littérature. Cependant, la modélisation de l'ensemble des jeux de policiers et voleurs ouvre la porte à une bonne quantité de problématiques intéressantes qui peuvent maintenant être abordées d'un niveau global tel qu'il n'était pas possible avant. Ainsi, bien qu'il est probable que notre modèle ne permette généralement pas d'obtenir de solutions analytiques à des questions classiques telles que la détermination du nombre minimal de policiers requis pour capturer un voleur, nous avons bon espoir que de belles solutions algorithmiques pourront être conçues pour résoudre ces mêmes questions sur des classes plus larges, non pas de graphes, mais de jeux. Par exemple, on peut maintenant demander sous quelles conditions sur la vitesse du voleur et la zone de surveillance du policier un seul policier peut gagner sur un graphe. On peut s'interroger sur l'indice n auquel on observe un changement de phase dans la récurrence (3.3). Autrement dit, il peut exister un nombre de tours n tel que la probabilité que le policier gagne en au plus n tours est très faible alors que la probabilité qu'il gagne en plus de n tours serait très forte.

Bref, il semble que de nouvelles et prometteuses avenues de recherche soient mises au jour avec les objets présentés dans ce chapitre.

Chapitre 4

Applications des jeux de poursuite en recherche opérationnelle

4.1 Introduction

En théorie de la recherche, branche de la recherche opérationnelle, est étudié le problème classique du *Optimal Search Path* (OSP). Le but de ce problème est de déterminer le chemin de longueur finie maximisant la probabilité qu'un évadé invisible errant sur un graphe se fasse capturer par un seul chercheur. Cette description justifie bien son nom anglais. Dans ce mémoire, on lui accorde cependant le nom français du *jeu du perdu* vu sa proximité avec les jeux de poursuite et qu'il est naturel de modéliser le comportement d'une personne perdue comme se déplaçant aléatoirement.

Notre contribution principale dans ce chapitre est d'avoir reconnu la proximité de ce jeu avec une variante du jeu de policier et voleur et d'avoir su utiliser celle-ci pour concevoir une borne supérieure sur la fonction objectif de ce problème. On montre en effet que la probabilité de capture du perdu peut servir dans une méthode de *branch and bound*.

Dans cette variante du jeu de policier et voleur que l'on nomme le jeu du *policier et voleur saoul* [25, 28] le voleur se déplace aléatoirement sur le graphe en pigeant à chaque tour un sommet au hasard parmi l'ensemble de ses sommets adjacents (incluant sa position). Autrement dit, il réalise une marche aléatoire. De plus, lorsque les deux joueurs se rencontrent sur un sommet, le policier n'est jamais assuré de mettre la main au collet du voleur, cet événement se réalise plutôt avec une probabilité connue a priori du policier. Autrement, ce jeu est en tout point identique au jeu de policier et voleur de Nowakowski et Winkler [39].

Remarque 4.1.1. *Ce chapitre découle d'un article présenté à la conférence The 20th International Conference on Principles and Practice of Constraint Programming (CP 2014) à Lyon, France, dans le cadre du programme doctoral. Une version longue, et améliorée, de cet article*

a été présenté à l'édition 2015 de cette conférence annuelle CP 2015 à Cork, Irlande [52].

4.2 Théorie de la recherche et OSP

La théorie de la recherche peut être simplement décrite à l'aide de son exemple d'application typique : la recherche d'objets (navires, avions) perdus en mer. Différentes instances de problèmes de ce domaine se décrivent en modifiant différentes contraintes : on peut décrire le problème en temps continu, chercher un objet dans un espace continu, avoir pour mandat de capturer plusieurs objets, etc. La particularité des jeux de perdus est le manque d'information des chercheurs en ce qui a trait à la position du (ou des) perdu(s). En effet, le jeu se déroule à information imparfaite, voire nulle, les chercheurs doivent donc composer avec le fait de ne pas pouvoir toujours calculer leur stratégie optimale puisque la position de l'objet cherché n'est pas connue.

Le problème de l'OSP est une instance très simple que l'on retrouve dans cette branche de la recherche opérationnelle. En effet, la structure du problème est complètement discrétisée, la vitesse des joueurs est ramenée à son minimum et même le nombre de joueurs est fixé à son plus bas. Ce problème est donc probablement l'instance la plus simple de la théorie de la recherche. Toutefois, malgré cette simplicité il reste que la fonction objectif de ce jeu est NP-difficile¹ à maximiser [54]. Bref, bien que le problème de l'OSP soit une simplification d'un problème plus général, celui-ci est fondamentalement difficile. Ceci le rend aussi intéressant.

4.3 Le problème de l'OSP, ou le jeu du perdu

Le jeu du perdu modélise une opération de recherche simplifiée se déroulant sur un graphe en temps discret. Un seul chercheur tente de retrouver un objet se déplaçant aléatoirement de sommet en sommet voisin sans jamais divulguer sa position au chercheur. La marche aléatoire de l'objet est cependant connue du chercheur. Le chercheur désire déterminer le chemin, fini, sur le graphe maximisant la probabilité qu'il capture l'objet. On précise que le chercheur ne connaît que des distributions de probabilités décrivant la position de l'objet : sa marche aléatoire, sa distribution initiale ainsi que les probabilités qu'il soit capturé.

L'exemple 4.3.1 illustre le problème du perdu sur une grille carrée.

Exemple 4.3.1. Supposons que le jeu se déroule sur une grille. Un seul perdu débute le jeu avec probabilité 1 sur le sommet à une des extrémités du graphe (le sommet en rouge de la figure 4.1(a)). La probabilité d'apparition du perdu sur les sommets du graphe au début du jeu est dictée par sa distribution initiale qui est connue du chercheur. L'unique chercheur

1. Pour plus d'information sur la théorie de la complexité et, en particulier, la classe des problèmes NP-complet, on renvoie au livre de Wegener [58]. On fait un abus de langage en disant que la fonction objectif de l'OSP est NP-difficile à maximiser puisque c'est le problème de décision qui lui correspond qui l'est.

choisit ensuite sa position de départ en se positionnant sur le centre du graphe (au sommet S_1). Après ce mouvement le perdu choisit une distribution uniforme sur son voisinage fermé, par exemple une distribution uniforme tel qu'illustré sur la figure 4.1(c). Les probabilités de transition sont contenues dans la matrice M de mouvements du perdu et sont connues du chercheur. Finalement, si le chercheur se déplace en S_2 , alors avec une certaine probabilité il y trouvera le perdu, autrement il pourra déduire que celui-ci se retrouve sur les deux sommets rouges restants de la figure 4.1(c) avec probabilité $\frac{1}{2}$.

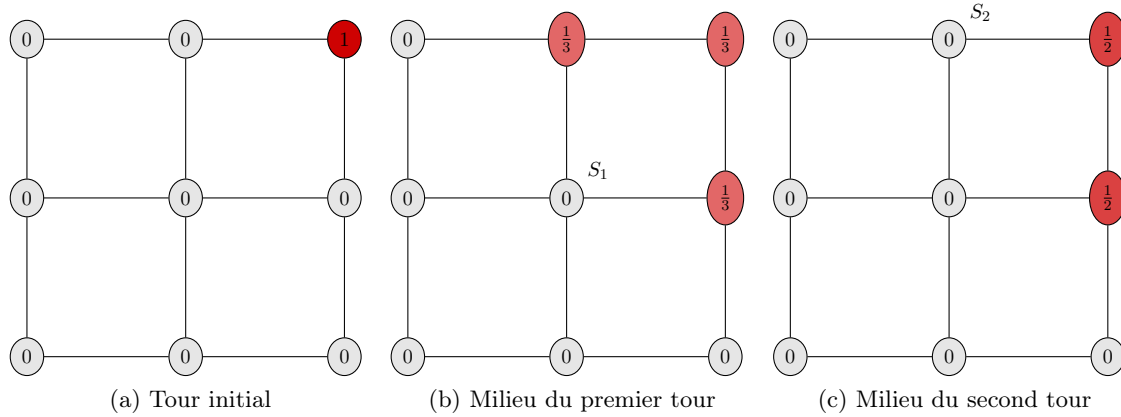


FIGURE 4.1 – Le jeu du perdu sur une grille

Soit un graphe $G = (V, E)$ que l'on suppose fini, réflexif et connexe. La finitude de G est importante, puisqu'il semble absurde de lancer une opération de recherche sur un graphe infini et que de toute manière on contraint le jeu à se terminer en temps fini. De même la connexité de G est naturelle, autrement il serait possible que le chercheur se trouve sur une composante connexe distincte de celle sur laquelle se trouverait l'objet et ce dernier ne serait jamais découvert. Finalement, G doit être réflexif (il y a une boucle autour de chaque sommet) pour permettre à chaque joueur de rester sur place. Il sera plus tard supposé que le graphe n'est pas dirigé, mais le modèle peut être aisément étendu au cas de graphes dirigés. On renvoie le lecteur au chapitre 1 pour des clarifications sur la théorie des graphes et des marches aléatoires. Dans le texte subséquent, une partie de la notation utilisée ne suit pas la notation présentée dans le chapitre précédent. Nous avons préféré réutiliser le plus possible les mêmes notations que celles qui se retrouvent dans la littérature sur le problème OSP. Toutefois, il apparaissait qu'une partie de cette notation n'était pas assez précise pour les fins de ce mémoire, auquel cas des précisions ont été apportées.

Deux joueurs, un chercheur et un perdu, se promènent sur G , le chercheur tentant de retrouver le perdu. Le perdu est invisible au chercheur et se déplace selon une marche aléatoire dictée par une matrice de transition M . Les deux joueurs amorcent chaque partie sur deux sommets quelconques de V ou d'un sous-ensemble $X \subseteq V$. Un maximum de T tours est alloué au

chercheur pour la capture du perdu. On considère qu'il existe pour chaque sommet $v \in V$ une variable aléatoire $D(v) \in \{0, 1\}$ telle que :

$$D(v) = 1 \iff \begin{array}{l} \text{Lorsque le chercheur et le perdu sont sur } v, \text{ le perdu est détecté} \\ \text{par le chercheur.} \end{array}$$

Alors, chaque fois que le chercheur et le perdu se rencontrent sur un sommet $v \in V$, il y a une probabilité

$$\text{pod}(v) := \mathbb{P}[D(v) = 1 \mid c = o = v] \quad (4.1)$$

que le chercheur détecte le perdu sur le sommet v ². Ici, c et o dénotent respectivement les positions courantes du chercheur et de l'objet. On pose $\text{pod}(v) = 0$ lorsque $c \neq o$ et $c = v$, soit lorsque le chercheur se tient sur v , mais que l'objet n'y est pas. Dans ce cas, aucune fouille n'est produite et cette probabilité ne devrait pas être définie, mais il est utile qu'elle le soit dans le reste du modèle. Les variables $c \in V$ et $o \in V$ peuvent être indicées par un entier t pour dénoter la position au temps t . On suppose que lorsque le perdu est détecté sur un sommet v , le jeu se termine et il est impossible qu'il soit détecté à nouveau ailleurs. On suppose donc qu'il existe un élément $j^* \notin V$ où est envoyé le perdu lorsqu'il est capturé. j^* est une technicité pour permettre de bien définir la probabilité $\text{pod}(v)$ et n'a pas été implémentée en pratique dans les expérimentations qui suivent.

Comme le chercheur connaît la probabilité $\text{pod}(v)$ ainsi que le modèle de mouvement du perdu, sous la forme de la matrice M , il peut calculer la probabilité que le perdu se retrouve sur un sommet v à n'importe quel incrément de temps. Ainsi, soit $\pi_t = (v_1, v_2, \dots, v_t)$ un chemin de $t \leq T$ sommets de G . On identifie toujours le mouvement du chercheur à un chemin de longueur au plus T . On note

$$\text{poc}_{\pi_t}(v) := \mathbb{P}[o_t = v \mid \pi_t = (c_1, c_2, \dots, c_t), (D(c_1), D(c_2), \dots, D(c_t)) = \mathbf{0}] \quad (4.2)$$

comme la probabilité que l'objet recherché se retrouve sur le sommet v au temps t tout en sachant que le chercheur a emprunté le chemin π_t et que l'objet n'a encore jamais été découvert³. On note que l'on utilise la notation vectorielle $\mathbf{0} = (0, 0, \dots, 0)$ pour simplifier certaines équations. À ces deux probabilités, on en ajoute une troisième qui permettra d'écrire la fonction objectif, soit

$$\text{cos}_{\pi_t} := \mathbb{P}[\exists! v \in \pi_t : D(v) = 1 \mid \pi_t = (c_1, c_2, \dots, c_t)] \quad (4.3)$$

$$= \sum_{i=1}^t \text{poc}_{\pi_i}(c_i) \text{pod}(c_i). \quad (4.4)$$

2. pod vient de l'anglais *probability of detection*, la probabilité que l'objet soit détecté.

3. La notation poc vient de l'anglais *probability of containment* que l'on traduit comme la probabilité qu'un sommet contienne l'objet.

Ainsi, \cos_{π_t} ⁴ est la probabilité que l'objet soit détecté quelque part sur le chemin π_t emprunté par le chercheur. La condition *il existe un unique* ($\exists!$) est un peu redondante puisque l'objet ne peut être capturé plus d'une fois. Il est bien cependant de le préciser.

Finalement, l'objectif du problème du perdu peut être décrit en terme de la probabilité \cos_{π_t} . Le chercheur tente de déterminer le chemin π_T^* maximisant la probabilité de capturer le perdu en au plus T incréments de temps. Alors, en posant \mathcal{P}_T l'ensemble des chemins de longueur T du graphe G , le chercheur doit calculer

$$\pi_T^* := \operatorname{argmax}_{\pi_T \in \mathcal{P}_T} \cos_{\pi_T}.$$

On utilise aussi la notation $\cos_{\pi_t}^*$ pour dénoter la probabilité maximale de capture du perdu avant la fin du jeu lorsque le chercheur a déjà parcouru le chemin π_t , soit

$$\begin{aligned} \cos_{\pi_t}^* &:= \max_{\pi_{T-t+1}} \mathbb{P} [\exists! v \in \pi_t \cup \pi_{T-t+1} : D(v) = 1 \mid \pi_t = (c_1, \dots, c_t)] \\ &= \cos_{\pi_t} + \max_{\pi_{T-t+1}} \mathbb{P} [\exists! v \in \pi_{T-t+1} : D(v) = 1] \end{aligned}$$

Cette définition, ainsi que l'égalité qui suit, sera importante dans la section 4.7 lorsque la borne sera définie. Effectivement, on utilisera le fait que cette probabilité soit décomposable de la sorte pour définir une borne supérieure sur la quantité $\max_{\pi_{T-t+1}} \mathbb{P} [\exists! v \in \pi_{T-t+1} : D(v) = 1]$.

Étant donné que le perdu se déplace aléatoirement selon une marche aléatoire gouvernée par une matrice M connue du chercheur, ce dernier peut toujours calculer la probabilité que le perdu transite d'un sommet à un autre après qu'une recherche ait été réalisée au premier sommet, soit

$$\begin{aligned} \operatorname{poc}_{\pi_t}(o') &= \operatorname{poc}_{\pi_t[t-1]}(v_t)(1 - \operatorname{pod}(v_t))M(v_t, o') \\ &\quad + \sum_{o \in V \setminus \{v_t\}} \operatorname{poc}_{\pi_t[t-1]}(o)M(o, o'), \quad \forall o' \in N[o]. \end{aligned} \tag{4.5}$$

Ici, $\pi_t[t-1]$ fait référence aux $t-1$ premiers sommets du chemin π_t . Autrement dit, l'équation (4.5) donne la probabilité que le perdu (ou l'objet) se trouve en position $o' \in N[o]$ alors que le chercheur a fouillé tous les sommets du chemin π_t . Il n'y a forcément que deux manières que cet événement se produise : soit l'objet se trouvait sur un sommet v_t fouillé par le chercheur au tour précédent (avec probabilité $\operatorname{poc}_{\pi_t[t-1]}(v_t)$), ne s'est pas fait détecter lors de la fouille avec probabilité $1 - \operatorname{pod}(v_t)$ et a transité vers sa nouvelle position o' ; autrement, l'objet ne partageait pas la position du chercheur et pouvait donc se trouver sur n'importe quel sommet o avec probabilité $\operatorname{poc}_{\pi_t[t-1]}(o)$ duquel il s'est dirigé vers o' .

4. On utilise la notation \cos venant de l'anglais *cumulative probability of success* en définissant un succès comme une capture de l'objet par le chercheur.

4.4 L’OSP en programmation par contraintes (CP)

4.4.1 Modélisation en CP

Dans notre article présenté à la conférence CP 2015 [52] est exposé à la section 2 un modèle du problème OSP sous la forme d’un programme par contraintes. Ce programme réutilise la définition mathématique de l’OSP décrite similairement à la section précédente, mais en définissant rigoureusement les variables de décision, leurs domaines respectifs ainsi que chaque contrainte explicite et implicite. Ce programme fut originellement décrit dans un article de Michael Morin et coll. [37] en 2012, puis réutilisé dans l’article accepté et présenté à la conférence CP 2015. Quoique ce modèle ne soit pas l’œuvre originale du présent auteur, étant donné son importance pour la suite celui-ci est retranscrit ici. On renvoie le lecteur intéressé à la section 1.3 pour de plus amples informations sur l’optimisation par contraintes (CP).

Le modèle CP du problème du perdu se décrit directement en partant de la définition du jeu. Celui-ci contient T variables. Pour chaque incrément de temps $t \in \{1, \dots, T\}$, on définit une variable de décision $PATH_t$ représentant la position du chercheur au temps t . Le nom de la variable renvoie au fait qu’en prenant l’ensemble des variables $PATH_t$ pour $t \in \{1, \dots, T\}$ on obtient un chemin. Le domaine de cette variable est un domaine énuméré égal à l’ensemble des sommets du graphe G , c’est-à-dire $\text{dom}(PATH_t) = V$. La position initiale du chercheur contraint le domaine de la variable $PATH_1$, ce qui restreint aussi les domaines de l’ensemble des variables de chemin. Ces domaines sont aussi limités par la contrainte triviale $[PATH_{t-1}, PATH_t] \in E, \forall t \in \{1, \dots, T\}$ spécifiant que le chercheur ne peut se déplacer qu’en suivant les arêtes du graphe. Cette contrainte est encodée avec la contrainte TABLE [4]. Le solveur doit donc déterminer une assignation aux variables $PATH_t, \forall t \in \{1, \dots, T\}$, correspondant aux T mouvements du chercheur sur le graphe sous la contrainte que l’assignation complète des variables $PATH_t$ forme un chemin. Cette assignation, qui représente un plan de recherche complet, doit maximiser la probabilité de détection de l’objet, comme on le verra plus loin.

Deux ensembles de variables implicites sont ensuite définis avec des domaines bornés pour définir les probabilités : $POC_t(v)$ ayant pour domaine $\text{dom}(POC_t(v)) = [0, 1]$ et $POS_t(v)$ ayant pour domaine $\text{dom}(POS_t(v)) = [0, 1]$ pour tout incrément de temps $t \in \{1, \dots, T\}$ et sommet $v \in V$. La variable POC_t représente la distribution de probabilité sur l’ensemble des sommets du graphe donnant la probabilité que le perdu se retrouve sur n’importe quel sommet $v \in V$ au temps t , soit $\text{poc}_{\pi_t}(v)$. Similairement, la variable POS_t est la distribution de probabilité donnant la probabilité que le perdu soit capturé au tour t . Bref, lorsque le chercheur se retrouve sur le sommet v au tour t , la probabilité que le perdu soit capturé est donnée par la contrainte

$$PATH_t = v \implies POS_t(v) = POC_t(v)\text{pod}(v). \quad (4.6)$$

Dans le modèle CP on note $\text{pod}(v)$ la probabilité de détection du perdu par le chercheur lorsque les deux partagent le même sommet, telle que définie à l'équation (4.1).

Si le chercheur ne se trouve pas sur le sommet v au tour t , alors la probabilité d'y détecter l'objet est clairement nulle. La contrainte suivante modélise ceci :

$$PATH_t \neq v \implies POS_t(v) = 0. \quad (4.7)$$

Les valeurs des variables POC_1 sont fixées en sachant la distribution initiale du perdu sur le graphe. Pour tout incrément de temps subséquent $t \in \{1, \dots, T\}$ et sommet $v \in V$, la connaissance que le chercheur possède sur la position du perdu évolue selon l'équation (4.8).

$$POC_t(v) = \sum_{r \in V} M(r, v) (POC_{t-1}(r) - POS_{t-1}(r)). \quad (4.8)$$

Une variable de domaine borné est ajoutée au modèle pour représenter la valeur objectif à maximiser, soit Z telle que $\text{dom}(Z) = [0, 1]$. La valeur objectif est la probabilité totale, par rapport aux positions possibles du perdu, de capturer le perdu, soit la valeur maximale sur l'ensemble des chemins de longueur T de cos_{π_T} . Comme π_T correspond à une assignation complète des variables $PATH_1, \dots, PATH_T$, la fonction objectif s'écrit finalement

$$Z = \sum_{t=1}^T \max_{v \in V} POS_t(v), \quad (4.9)$$

$$\max_{PATH_1, \dots, PATH_T} Z. \quad (4.10)$$

4.4.2 Le problème du perdu résolu par branch and bound

En théorie de la recherche et, plus particulièrement dans le problème de l'OSP, il est déjà commun d'utiliser une méthode de branch and bound pour résoudre l'OSP [57]. L'exemple qui suit illustre l'importance de développer des bornes supérieures performantes sur la fonction objectif du problème du perdu.

Exemple 4.4.1. Restons dans le contexte de la recherche de perdu et imaginons un bateau en mer pris dans un épais brouillard à la recherche du naufragé d'un navire survivant sur un radeau de fortune. Le temps pour ce naufragé est compté, il est affamé et frigorifié par les eaux. Bref, il doit rapidement être rescapé de son malheur. Le capitaine du bateau a en sa possession une carte maritime divisée en grille, formant un graphe, sur laquelle sont indiqués la position exacte de son bateau et les endroits où pourrait se trouver le naufragé. On sait que les courants marins font voguer le naufragé d'une case de la grille à l'heure. Ceci dit, s'il n'a au plus que 48 heures à vivre avant, par exemple, de souffrir d'hypothermie, quelle méthode est-ce que le capitaine du bateau devrait employer pour maximiser ses chances de le sauver à temps ?

Supposons que le capitaine, de par son expérience en mer, puisse garantir qu'un chemin π mène à une probabilité $p(\pi)$ que le naufragé soit rescapé à temps. Suivant les idées de Washburn [57], le capitaine pourrait énumérer chaque chemin π' et borner supérieurement la probabilité $p(\pi')$ par une probabilité $q(\pi')$, soit $p(\pi') < q(\pi')$. Lorsque cette dernière probabilité est plus faible que $p(\pi)$, c'est-à-dire $q(\pi') < p(\pi)$, alors π' ne peut être une solution intéressante puisqu'on a $p(\pi) > q(\pi') > p(\pi')$. Ainsi, dans le meilleur des scénarios, le chemin π' ne permet de sauver le naufragé qu'avec une probabilité inférieure à $p(\pi)$ qui, elle, est garantie puisque c'est une solution déjà calculée.

La vitesse à laquelle est calculée la borne supérieure ainsi que sa proximité avec la valeur exacte sont deux points majeurs à considérer dans le développement d'une telle borne. En effet, on veut généralement résoudre un problème NP-difficile, il est donc naturel de ne pas concevoir une borne NP-difficile, ou pire, à calculer. Dans tous les cas, on veut donc une borne qui se calcule en temps polynomial sur la taille du graphe d'un ordre le plus petit possible. D'un autre côté, dans le cas d'une probabilité, une borne supérieure triviale est toujours la valeur 1, ce qui n'est d'aucune utilité. Plus une borne supérieure est serrée, plus elle permet de retirer des solutions de l'ensemble des solutions possibles. Plus une borne est près de la valeur exacte, plus elle sera lente à calculer. La question difficile est donc de trouver le compromis idéal entre la précision d'une borne et sa rapidité de calcul.

4.5 La borne DMEAN

Il existe déjà dans la littérature plusieurs bornes supérieures sur la fonction objectif de l'OSP permettant d'appliquer un branch and bound. Notamment, Martins [36] présente la borne MEAN pour borner la fonction objectif. Nous nous concentrons ici sur la borne présentée par Lau et coll. [30], DMEAN, qui resserre la borne de Martins tout en s'assurant qu'elle reste valide. Dans cette section est présentée en premier lieu la borne de Martins, MEAN, puis celle de Lau et coll., DMEAN. Les deux sont reformulées dans notre modèle CP pour l'OSP.

Le choix d'implémenter la borne DMEAN au lieu d'une autre borne est justifié d'un côté parce que celle-ci est relativement récente et performe bien dans l'OSP et d'un autre côté parce qu'autant celle-ci que notre borne COPWIN sont basées sur des concepts de théorie des graphes. DMEAN borne la fonction objectif en résolvant un problème du plus long chemin dans un graphe dirigé acyclique (DAG) judicieusement construit alors que COPWIN, tel que précisé plus loin, est basée sur les jeux de poursuite et utilise, à notre opinion, une relaxation plus naturelle.

MEAN, autant que son extension DMEAN, cherche à maximiser le nombre espéré de détections réalisées à partir d'un état de jeu précis jusqu'à la fin du jeu. L'idée centrale de MEAN est la construction d'un DAG pour chaque incrément de temps $t \in \{1, \dots, T\}$ avec comme ensemble

de sommets des paires (r, k) telles que $r \in V$ et $t < k \leq T$. Une arête dirigée entre les sommets (r, k) et $(r', k + 1)$ est placée si et seulement si $[r, r'] \in E$, indiquant qu'il est possible pour les joueurs de passer de la position r au temps k à la position r' au tour suivant.

La borne DMEAN borne la probabilité de capture du perdu en comptant le nombre maximal espéré de détections dans le jeu. Selon les règles, ce nombre ne peut dépasser 1, donc l'espérance du nombre de captures est égale à la probabilité qu'il y ait une capture. La borne DMEAN permet de relâcher cette règle et permettre plusieurs captures. Ainsi, en comptant le nombre espéré de captures selon DMEAN on obtient une valeur plus grande que le nombre réel de captures selon les règles, soit la fonction objectif.

L'espérance maximale du nombre de détections Z_{DAG} du tour t au tour T est obtenue suivant le chemin le plus long sur le DAG, ce qui peut être calculé en temps $|V| + |E|$. Étant donné un chemin π_t , il est suffisant de connaître la distribution poc_{π_t} pour pouvoir calculer la probabilité de détection⁵ au tour k , avec $t \leq k \leq T$, et au sommet $r \in V$ puisque celle-ci donne la probabilité que le perdu se retrouve sur n'importe quel sommet au temps t une fois que le chercheur a fouillé le chemin π_t . Un poids $R_{\text{MEAN}}(r', k + 1)$ est placé sur une arête entrante dans le sommet $(r', k + 1)$ (partant d'un sommet (r, k)) tel que

$$R_{\text{MEAN}}(r', k + 1) := \sum_{v \in V} \text{poc}_{\pi_t}(v) M^{k+1-t}(v, r') \text{pod}(r'). \quad (4.11)$$

t est donc le nombre de tours joués, alors que les poids sur le DAG sont ajoutés pour tout $k > t$. Ici, M^{k+1-t} est la matrice de transition M à la puissance $k + 1 - t$ qui contient les probabilités de transitions en $k + 1 - t$ tours. Autrement dit, $M^{k+1-t}(v, r')$ donne la probabilité que le perdu transite du sommet v au sommet r' en $k + 1 - t$ tours. Ainsi, $R_{\text{MEAN}}(r', k + 1)$ est la probabilité que le perdu se retrouve sur un sommet quelconque v au tour t alors que le chercheur a fouillé le chemin π_t , marche encore $k + 1$ pas pour se retrouver sur le sommet r' et s'y fasse capturer par le chercheur. MEAN est une borne supérieure puisqu'on ne considère pas les fouilles que le chercheur peut avoir réalisées sur le graphe entre les tours t et $k + 1 - t$. MEAN pose donc une probabilité uniforme sur les arêtes $[(r, k), (r', k + 1)]$ pour tous les sommets (r, k) .

La borne DMEAN est construite à partir de l'idée qu'un perdu transitant du sommet r vers le sommet r' doit avoir survécu à une détection sur le sommet r pour possiblement se faire capturé sur le sommet r' . Ainsi, sur chaque arête sortante d'un sommet (r, k) vers un sommet $(r', k + 1)$ on place le poids $R_{\text{DMEAN}}(r', k + 1)$ défini par :

$$R_{\text{DMEAN}}(r', k + 1) = \begin{cases} R_{\text{MEAN}}(r', k + 1), & \text{si } k = t; \\ R_{\text{MEAN}}(r', k + 1) \\ -R_{\text{MEAN}}(r, k)M(r, r')\text{pod}(r'), & \text{autrement.} \end{cases} \quad (4.12)$$

Si $t = k$, alors $R_{\text{MEAN}}(r', k + 1) = \sum_{v \in V} \text{poc}_{\pi_t}(v)M(v, r')\text{pod}(r')$, soit la probabilité que le perdu soit présent, sur le graphe, transite directement vers r' et s'y fasse détecter. Ainsi,

le perdu a probabilité $R_{\text{MEAN}}(r', k + 1)$ d'atteindre le sommet r' au temps $k + 1$ et de se faire détecter sur ce sommet. Si $k > t$, on soustrait de cette probabilité la probabilité $R_{\text{MEAN}}(r, k)M(r, r')\text{pod}(r')$. Ce dernier terme est la probabilité que le perdu se soit fait capturer en r , transite vers r' et s'y fasse aussi capturer. Comme cette probabilité est soustraite, $R_{\text{DMEAN}}(r', k + 1)$ est la probabilité que le perdu survive à une fouille sur r , mais se fasse capturer en arrivant en r' . On a en effet, lorsque $k > t$,

$$\begin{aligned}
R_{\text{MEAN}}(r', k + 1) &:= \sum_{v \in V} \text{poc}_{\pi_t}(v) M^{k+1-t}(v, r') \text{pod}(r'), \\
\implies R_{\text{DMEAN}}(r', k + 1) &= \sum_{v \in V} \text{poc}_{\pi_t}(v) M^{k+1-t}(v, r') \text{pod}(r') \\
&\quad - \text{pod}(r') \sum_{u \in V} \text{poc}_{\pi_t}(u) M^{k-t}(u, r) M(r, r') \text{pod}(r) \\
&= \text{pod}(r') \mathbb{P}[o_{t+k+1} = r' \mid \pi_t, D(\pi_t) = \mathbf{0}] \\
&\quad - \text{pod}(r') \mathbb{P}[o_{t+k} = r \mid \pi_t, D(\pi_t) = \mathbf{0}] M(r, r') \text{pod}(r).
\end{aligned}$$

Dans les dernières égalités, on a utilisé une version simplifiée de la définition de $\text{poc}_{\pi_t}(v)$ donnée par l'équation (4.2). Ainsi, on voit que la définition secondaire de DMEAN, le cas où $k > t$, donne la probabilité que le perdu se soit retrouvé sur le sommet r au tour $t + k$, y ait survécu une détection pour transiter vers le sommet r' et s'y faire prendre par le chercheur. Évidemment, le tout est conditionnel au fait que le chercheur ait parcouru le chemin π_t . On obtient donc la probabilité que le perdu se trouve en r' au tour $t + k + 1$ et s'y fasse capturer tout en sachant qu'il a survécu à une détection sur le sommet r au tour précédent. Finalement, contrairement à la borne MEAN qui a un caractère universel puisqu'elle applique le même poids à chaque arête $[(r, k), (r', k + 1)]$, la borne DMEAN module ce poids selon le sommet de départ. Ceci la rend plus serrée que la première borne.

Une borne est dite *admissible* si elle ne sous-estime (surestime) jamais la valeur objectif à maximiser (minimiser). Dans le problème de l'OSP, une borne admissible étant donné le début d'un chemin π_t est obtenue en calculant la valeur optimale Z_{DAG}^* d'un plus long chemin sur le DAG et en sommant ensuite Z_{DAG}^* et \cos_{π_t} [30].

Remarque 4.5.1. *Pour le calcul de la borne DMEAN, il est nécessaire de construire un graphe dirigé acyclique avec nT sommets et mT arêtes, lorsque l'OSP est résolu sur un graphe de n sommets et m arêtes. Ainsi, au tour k , la borne demande de résoudre un problème de plus long chemin sur ce nouveau graphe dirigé, nécessitant $O((T - k + 1)(n + m))$ étapes.*

5. Dans notre article de 2015 [52], on utilise parfois les termes *expected detection*. Ceci suggère que le terme *detection* fait référence à un événement aléatoire suivant une loi de Bernoulli dont la probabilité est la probabilité de détection. On parle aussi du nombre espéré de détections, ce qui est bien défini à partir de la suite de variables de Bernoulli définies comme la suite de détections et non-détections. Ces variables et probabilités sont donc bien définies et on remplace ici le terme *expected detection* par l'expression plus naturelle *probabilité de détection*.

Pour appliquer la borne DMEAN dans notre modèle CP, il est justifié d'ordonner les variables de chemin selon la position du chercheur dans le temps en suivant un ordre statique. Il est naturel de faire cela puisque les variables de décision sont déterminées par le chemin emprunté par le chercheur et que cet ordre décrit l'ordre dans lequel le problème se déroule. La borne DMEAN est démontrée admissible [30], ainsi chaque fois que le chercheur entre dans un sommet et que la valeur de DMEAN est plus petite que la borne inférieure de la valeur objectif, ce sommet est retiré de l'arbre de recherche pour le chercheur puisque cette position est démontrée comme inutile.

4.6 Le jeu du policier et du voleur saoul et évasif

Cette section présente une variante du jeu classique du policier et voleur dans lequel le voleur se promène de manière aléatoire sur le graphe et a une probabilité positive de s'évader du policier lorsque les deux joueurs se croisent sur un sommet.

4.6.1 Le jeu du policier et du voleur saoul

Ce jeu se déroule sur un graphe G fini et réflexif. On suppose encore que G est connexe pour simplifier le problème, mais cette hypothèse n'est pas fondamentale. Un seul policier poursuit un seul voleur sur G selon les mêmes règles que le jeu de policier et voleur classique (déplacement sur les voisins adjacents, information parfaite, etc.). La seule différence ici est que le voleur est contraint de se déplacer selon une matrice de transition (matrice de probabilités) M connue du policier. Bref, le voleur fait une marche aléatoire sur G et le policier tente de l'interrompre.

Pour illustrer le jeu de policier et voleur saoul, on reprend le graphe de l'exemple 4.3.1 et on y fait jouer ce nouveau jeu.

Exemple 4.6.1. Le jeu se déroule encore une fois sur une grille. Un seul voleur s'y promène en commençant la partie sur le sommet R_0 de la figure 4.2(a). Le policier débute en C_0 . Encore une fois, le voleur se déplace aléatoirement et, dans ce cas-ci, uniformément sur l'ensemble des sommets de son voisinage fermé. Autrement dit, il choisit de bouger vers n'importe quel sommet accessible avec probabilité $\frac{1}{3}$, ce que le policier remarque après s'être déplacé en C_1 tel que sur la figure 4.2(b). La différence ici avec le problème du perdu apparaît lorsque le voleur se déplace sur la position R_1 , ce que le policier remarque et peut utiliser comme information pour rejoindre le voleur sur la même position au tour suivant (voir la figure 4.2(c)).

4.6.2 La récurrence w_n

Tel que déjà présenté au chapitre 2, la méthode classique pour résoudre les jeux de policiers et voleurs consiste à définir une relation d'ordre récursive entre les sommets du graphe (ou une transformation de celui-ci), de démontrer que cette relation devient stationnaire puis de

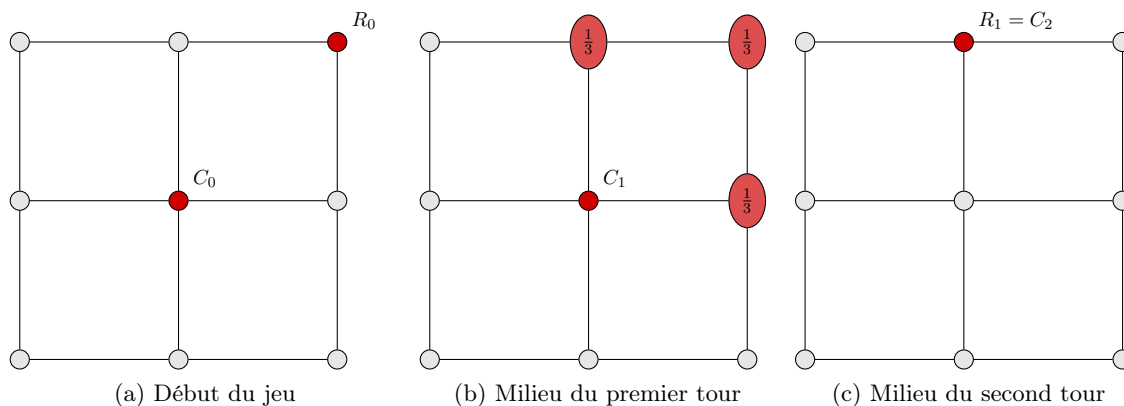


FIGURE 4.2 – Exemple de jeu de policier et voleur saoul.

déterminer si celle-ci est totale ou non. En effet, on démontre habituellement qu’il existe une stratégie gagnante pour le policier (au sens de la théorie des jeux) si et seulement si cette relation est totale. Bref, cette relation contient toute l’information nécessaire pour comprendre un jeu de policiers et voleurs typique.

Initialement apparu dans la littérature aux environs de 2012 dans une publication de Pralat et Kehagias [25] et une autre de Komarov et Winkler [27], le jeu de policier et voleur saoul est la première variante connue des jeux de policiers et voleurs intégrant des aspects stochastiques. Pralat et Kehagias, cependant, ne présentaient pas ce jeu dans l’optique de le résoudre, mais plutôt dans celle d’étudier le temps de capture moyen du voleur. Étonnamment, ceux-ci démontrent implicitement que le calcul du temps de capture moyen sur certains graphes ne nécessite pas de connaître la stratégie optimale du policier. Ceci est dû au fait qu’ils contraignent la marche du voleur à être uniforme (celui-ci ayant autant de chances de traverser n’importe quelle arête à partir de son sommet courant) et n’étudient que certains graphes simples (chemins, cycles, arbres, etc.)⁶. Quant à eux, Komarov et Winkler définissent aussi le jeu du voleur saoul avec l’objectif d’étudier son temps de capture. Leur principale contribution est une borne sur l’espérance du temps de capture sur l’ensemble des graphes à n sommets. C’est une fois de plus un résultat intéressant, mais qui ne nous éclaire pas sur la stratégie optimale du policier.

Ainsi, nous présentons à notre connaissance la première méthode générale pour résoudre le jeu du policier et du voleur saoul, peu importe le graphe sur lequel celui-ci est joué ainsi que la forme de la marche réalisée par le voleur.

Nous définissons une relation de récurrence $w_n(x, y)$ donnant la probabilité qu’un policier en y capture un voleur en x en au plus n tours. Ceci n’est donc pas directement une relation

6. Malgré ces restrictions, leurs résultats ne sont pas triviaux et nécessitent des objets venant des probabilités avancées tels que la borne de Chernoff, que nous ne présentons pas ici vu qu’elle ne sera pas utilisée.

d'ordre, il est d'ailleurs impossible de définir ce type de relation puisque le jeu est stochastique. Cependant, comme on peut voir la relation $x \preceq_n y : V^2 \rightarrow \{0, 1\}$ en tant que relation vers l'ensemble $\{0, 1\}$, on peut généraliser cette notion en définissant une autre relation $w_n(x, y) : V^2 \rightarrow [0, 1]$. Cette relation généralise donc la première, non pas en terme d'ordre, mais d'un point de vue numérique. De ce point de vue, on peut interpréter la relation $w_n(x, y)$ comme une probabilité.

Bref, on suppose qu'un policier et un voleur se promènent sur un graphe G . Le voleur se déplace selon une marche aléatoire sur G contrôlée par la matrice M , c'est-à-dire chaque élément (r, r') dans M donne la probabilité que le voleur traverse l'arête $[r, r'] \in E$. Le jeu se termine si et seulement si les deux joueurs se croisent sur un sommet. La proposition 4.6.3 démontre que la relation $w_n^M(r, c)$ donne la probabilité

$$w_n^M(r, c) := \mathbb{P}[T \leq n \mid r, c],$$

où T est le temps de capture du voleur par le policier lorsque le policier joue optimalement et que le voleur suit M .

Définition 4.6.2. Soit r, c les positions du voleur et du policier sur le graphe G et M la matrice de transition du voleur. Posons

$$w_0^M(r, c) := \begin{cases} 1, & \text{si } r = c; \\ 0, & \text{autrement.} \end{cases}$$

$$w_n^M(r, c) := \begin{cases} 1, & \text{si } c \in N[r], n \geq 1; \\ \max_{c' \in N[c]} \sum_{r' \in N[r]} M(r, r') w_{n-1}^M(r', c'), & \text{si } c \notin N[r], n \geq 1. \end{cases} \quad (4.13)$$

L'équation (4.13) est justifiée par la proposition suivante qui démontre que celle-ci donne bel et bien la probabilité recherchée.

Proposition 4.6.3. Lorsque le policier joue en premier sur G et que M contrôle la marche aléatoire du voleur, alors $w_n^M(r, c)$ donne la probabilité que le policier débutant en c capture le voleur débutant en r en au plus n tours.

Démonstration. Le résultat est démontré par induction sur n .

$n = 0$: Le cas de base $n = 0$ est clair puisque le policier gagne sans mouvement si et seulement s'il débute son tour sur la position du voleur.

$n > 0$: Autrement, supposons que la proposition tienne pour $n - 1 \geq 0$ et démontrons-le pour n . Si $c \in N[r]$, alors $w_n^M(r, c) = 1$ et le résultat suit du fait que le policier capture effectivement le voleur (on se rappelle que le policier joue en premier). Sinon, si $c \notin N[r]$, alors laissons le policier se déplacer vers un sommet $c' \in N[c]$. La position du voleur à

la fin du tour sera r' avec probabilité $M(r, r')$. Le policier doit moyenner la probabilité qu'il capture le voleur sur la prochaine position du voleur, autrement dit calculer

$$\mathbb{P}[T \leq n \mid r, c'] = \sum_{r' \in N[r]} \mathbb{P}[r' \mid r] \mathbb{P}[T \leq n-1 \mid r', c'] = \sum_{r' \in N[r]} M(r, r') w_{n-1}^M(r', c').$$

L'hypothèse d'induction a été utilisée pour écrire $\mathbb{P}[T \leq n-1 \mid r', c'] = w_{n-1}^M(r', c')$. Ainsi, la meilleure position à choisir pour le policier est donnée par

$$c^* = \operatorname{argmax}_{c' \in N[c]} \sum_{r' \in N[r]} M(r, r') w_{n-1}^M(r', c').$$

La probabilité recherchée est donc bien $w_n^M(r, c)$. □

À notre connaissance, la Proposition 4.6.3 est la première caractérisation de la stratégie optimale du policier dans le jeu de policier et voleur saoul. Celle-ci pourrait donner une borne triviale sur la probabilité de capturer le perdu dans le problème de l'OSP, mais une borne plus serrée est présentée à la section suivante. Cette prochaine borne utilise les processus de Markov décisionnels, ce qui facilite la démonstration de son admissibilité comme borne.

4.6.3 Le jeu du policier et voleur en processus de Markov décisionnel (MDP)

À la section 4.6.1 est présenté un jeu de policier et voleur dans lequel le voleur se déplace aléatoirement sur le graphe. Ce jeu est formulé dans ce qui suit sous la forme d'un processus de Markov décisionnel (*Markov Decision Process*, MDP) pour faciliter sa comparaison avec la fonction objectif du problème de l'OSP.

4.6.4 Quelques concepts de MDP

Conceptuellement le MDP peut être vu de deux manières. Typiquement, on le présente comme une généralisation de la chaîne de Markov dans laquelle un agent peut prendre des actions pour contrôler le processus se déroulant sur la chaîne, le terme *contrôle* est ici employé du fait qu'historiquement le MDP a une grande proximité avec la théorie du contrôle [45]. D'un autre côté, le MDP peut être vu comme un jeu répété à un seul joueur avec des transitions aléatoires entre les états. On reviendra sur ces deux points de vue dans quelques instants.

Définition 4.6.4. *Un MDP [45] $\mathcal{M} = (S, A, P, R)$ se décrit à l'aide d'un espace d'états dénombrable S , d'un ensemble d'actions dénombrable A , d'une probabilité de transition $P : S \times A \times S \rightarrow [0, 1]$ et d'une fonction de gains $R : S \times S \rightarrow \mathbb{R}$. À chaque incrément de temps, le système est dans un état $s \in S$. Un agent, ou contrôleur, doit choisir une action $a \in A$, auquel cas le prochain état s' du système est choisi aléatoirement selon la fonction $P(s, a, s')$*

interprétée comme la probabilité $\mathbb{P}[s' | s, a]$. Finalement, l'agent reçoit un gain réel $R(s, s')$ et recommence le processus à partir de l'état s' . L'objectif de l'agent est de maximiser l'espérance de son gain en un temps fini T .⁷

L'agent doit déterminer une *politique*, ou stratégie, soit une fonction $\sigma : S \rightarrow A$ associant à chaque état du processus une action à jouer. Chaque politique détermine entièrement un jeu et permet de calculer l'espérance de gain que l'agent peut obtenir durant une partie, soit $\sum_{t=0}^T \mathbb{P}[s_{t+1} | \sigma(s_t), s_t] R(s_t, s_{t+1})$, où s_t est l'état rencontré au temps t .

On peut voir la proximité avec la chaîne de Markov classique en posant $P(s, a, s') := \mathbb{P}[s' | s]$, auquel cas le système évolue toujours indépendamment de l'action choisie par l'agent. Celui-ci est, en quelque sorte, *coincé* dans une chaîne de Markov. Comme l'action de l'agent n'affecte aucunement le système, le gain devient inutile et les chaînes de Markov se décrivent donc sans gain.

On remarque aussi le lien avec la théorie des jeux en notant le gain que l'agent réalise après chaque transition. En effet, la théorie des jeux décrit les interactions d'agents égoïstes interagissant entre eux et voulant maximiser leurs gains sous certaines règles préétablies. Mais on a ici un agent qui cherche à maximiser son gain selon des règles préétablies. On peut imaginer que cet agent s'oppose à *dame nature* et qu'après chaque action, celle-ci choisit aléatoirement une action inconnue de l'agent et que l'état résultant de celle-ci est le nouvel état de jeu.

4.6.5 Retour au MDP du policier et voleur saoul

La prochaine définition établit le MDP à partir duquel sera développée la borne supérieure COPWIN sur la probabilité de capture du perdu dans l'OSP.

Définition 4.6.5 (MDP du policier et voleur saoul). *Soit G un graphe, M une matrice de transition définie sur V et T le nombre maximal d'incrément de temps alloués. On définit $\mathcal{M} = (S, A, P, R)$ comme suit. L'agent est le policier et les transitions aléatoires sont principalement contrôlées par M .*

Un état $s \in S$ est un triplet (r, c, t) contenant les positions du voleur et du policier $r, c \in V$ ainsi que le temps courant $t \in \{1, \dots, T + 1\}$. Une position particulière pour le voleur est ajoutée, prison, d'où il ne peut sortir. L'ensemble des actions est simplement V , l'ensemble des sommets sur lesquels le policier se déplace (on restreint avec P l'action du policier à n'être choisie que parmi les sommets adjacents). D'une paire de positions du policier et du voleur à un certain temps, après le choix d'action du policier le prochain état est tiré aléatoirement avec probabilité P . Finalement, un gain de 1 est donné au policier lorsque le voleur transite

7. Les processus infinis existent aussi, mais nécessitent quelques définitions de plus. Ils ne sont pas exposés ici, puisqu'on ne considère que des processus finis dans ce chapitre.

vers la position *prison*, c'est-à-dire lorsqu'il est capturé. Bref, on a :

$$S := (V \cup \{\text{prison}\}) \times V \times \{1, \dots, T + 1\}, \quad (4.14)$$

$$A := V, \quad (4.15)$$

$$P[(r', c', t') \mid (r, c, t), a] := \begin{cases} 1, & \text{si } r = r' = \text{prison}; \\ \text{pod}(r), & \text{si } r = c', r' = \text{prison}; \\ (1 - \text{pod}(r))M(r, r'), & \text{si } r = c', r' \neq \text{prison}; \\ M(r, r'), & \text{si } r \notin \{c, c', \text{prison}\}; \\ 0, & \text{si } a \neq c' \text{ ou } c' \notin N[c] \text{ ou } t' \neq t + 1. \end{cases} \quad (4.16)$$

$$R((r', c', t') \mid (r, c, t), a) := \begin{cases} 1, & \text{si } r' = \text{prison} \neq r, t \leq T; \\ 0, & \text{sinon.} \end{cases} \quad (4.17)$$

Au début du jeu, le policier choisit sa position initiale c parmi un sous-ensemble X des sommets du graphe, puis la position initiale du voleur r est choisie aléatoirement parmi les sommets du graphe selon la distribution poc_{π_0} (on rappelle que cette distribution existe puisqu'on est dans le contexte de l'OSP). Ces deux choix créent un état $(r, c, 1)$ pour le MDP.

Le policier joue en premier à chaque tour. Si le policier rejoint la position du voleur ($r = c'$), alors avec probabilité $\text{pod}(r)$ le voleur est capturé et transite vers la prison ($r' = \text{prison}$), auquel cas il y reste pour le restant de la partie. Le policier reçoit, une seule fois, un gain d'un point après sa capture. Si la capture échoue, avec probabilité $1 - \text{pod}(r)$, ou si le policier ne s'est pas déplacé vers la position du voleur ($r \neq c'$), alors le voleur est libre de continuer à jouer selon la matrice M .

Les probabilités de transition (4.16) sont non-nulles seulement lorsque des actions valides sont choisies, c'est-à-dire le compteur de tours incrémente de 1 et l'action $a = c'$ est dans le voisinage fermé de c .

Dans le MDP \mathcal{M} , l'objectif du policier est de déterminer une politique maximisant son gain espéré avant la fin des T tours. Comme le policier ne peut obtenir au plus qu'un point dans une partie, l'espérance de son gain correspond exactement à la probabilité que le voleur soit capturé en au plus T tours. Une stratégie dans \mathcal{M} détermine la sélection de la position initiale du policier, puis correspond à une fonction $f : S \rightarrow A$ prescrivant à l'état courant une unique action à choisir, c'est-à-dire vers quel état transiter. On note que le policier a toujours accès à l'information concernant la position du voleur puisque celle-ci est incluse dans l'état courant, il peut ainsi prendre cette information en compte en élaborant sa stratégie.

On peut remarquer la proximité entre cette description d'un jeu de policier et voleur sous la forme d'un processus de Markov contrôlé par un seul agent et le modèle 3.2.1 décrivant

l'ensemble des jeux de policiers et voleurs. D'ailleurs, le MDP 4.6.5 offre une description très similaire à celle du modèle 3.2.1 pour ce jeu spécifique.

Comme un MDP satisfait la propriété de Markov, on peut aisément calculer la *valeur*⁸ $u_f(r, c, t)$ d'un état (r, c, t) du système lorsqu'une stratégie f est fixée. La valeur d'un état correspond au gain espéré que le policier peut obtenir en suivant la stratégie f à partir de l'état (r, c, t) . La valeur maximale u^* est donc celle maximisant la valeur d'un état parmi l'ensemble des stratégies possible d'un ensemble \mathcal{F} , soit $u^*(r, c, t) = \max_{f \in \mathcal{F}} u_f(r, c, t)$. La stratégie optimale du policier consiste à se déplacer vers la position du voleur lorsque possible et ensuite de tenter de le capturer. Si le voleur est hors de portée, le policier se dirige vers la position maximisant sa probabilité de gagner dans le temps restant à partir du prochain tour. Similairement à la proposition 4.6.3, la valeur de la stratégie optimale est donnée par :

$$u^*(r, c, t) = \begin{cases} \max_{c' \in N[c]} \sum_{r' \in N[r]} M(r, r') u^*(r', c', t + 1), & \text{si } r \notin N[c], t < T; \\ 1 - (1 - \text{pod}(r))^{T+1-t}, & \text{si } r \in N[c], t \leq T; \\ 0, & \text{si } r \notin N[c], t = T. \end{cases} \quad (4.18)$$

Si $r \notin N[c]$, alors le policier, jouant en premier, doit choisir sa prochaine position maximisant la probabilité qu'une éventuelle capture se produise en prenant en considération la position actuelle du voleur et le fait que le mouvement de celui-ci est contrôlé par M . Autrement, si $r \in N[c]$, le policier tente de capturer le voleur avec probabilité de capture $\text{pod}(r)$; s'il échoue, alors comme le voleur reste forcément dans le voisinage du policier ce dernier peut retenter de le capturer jusqu'à ce qu'il y ait capture ou que le jeu soit terminé. Cette formule utilise aussi le fait qu'au début de son tour, le policier a la possibilité de faire au plus $T + 1 - t$ tentatives de capture. Le jeu ici n'est pas simultané, il est donc impossible que les joueurs se croisent sur une arête sans qu'il y ait une tentative de capture. L'équation (4.18) est analogue à l'équation (4.13) avec les incréments de temps inversés et les probabilités de capture $\text{pod}(r) = 1$ pour tous les sommets r . L'intérêt de la récurrence (4.13) est sa simplicité. Cette récurrence est, à notre connaissance, la première formulation de la stratégie optimale du policier dans le jeu de policier et voleur saoul. La seconde équation, l'équation (4.18), est cependant préférée puisque'elle est plus près de la formulation de la fonction objectif du jeu de l'OSP.

4.7 La borne COPWIN

Conceptuellement, la borne COPWIN est probablement la borne la plus simple à concevoir. Malgré qu'elle se décrive mathématiquement, voire verbalement, de manière parfois obscure puisqu'elle nécessite l'emploi de deux jeux de poursuite différents chacun avec leurs vocabulaires propres, avec recul elle semble pratiquement triviale. Bref, comme il est nécessairement

8. Sans la propriété de Markov, il serait possible de calculer la valeur d'un état, mais hautement inefficace puisqu'il serait nécessaire d'utiliser l'historique complet des états parcourus. C'est entre autres ce qui justifie l'intérêt des processus markoviens.

	OSP	Policier-voleur saoul
Déroulement :	jeu alterné	jeu alterné
Nombre de joueurs :	1 chercheur vs 1 objet	1 policier vs 1 voleur
Structure :	graphe	graphe
Déplacement des joueurs :	sur les sommets	sur les sommets
Capture :	probabiliste	probabiliste
Temps de jeu :	au plus T tours	au plus T tours
Stratégies du chercheur / policier :	déterministe	déterministe
Stratégies de l'objet / du voleur :	marche aléatoire	marche aléatoire
Fonction objectif :	probabilité de capture en au plus T tours	probabilité de capture en au plus T tours
Information :	nulle	parfaite.

FIGURE 4.3 – Résumé des caractéristiques partagées et distinctives entre l'OSP et policier et voleur saoul

plus facile de résoudre le jeu du policier et voleur saoul que le problème de l'OSP, c'est-à-dire de capturer un voleur visible que de trouver un objet invisible, la probabilité de capture dans l'OSP est plus faible que la probabilité de capture dans le jeu de policier et voleur correspondant. Ceci est démontré dans la proposition qui suit. On résume avant les caractéristiques qui sont partagées et celles qui diffèrent entre les deux jeux de l'OSP et du policier et voleur saoul dans le tableau 4.7.

Proposition 4.7.1. *La probabilité $\cos_{\pi_0}^*$ de trouver le perdu dans le problème de l'OSP est toujours au plus celle de capturer le voleur,*

$$\cos_{\pi_0}^* \leq \max_{c \in X} \sum_{r \in V} \text{poc}_{\pi_0}(r) u^*(r, c, 1),$$

avec $X \subseteq V$ le sous-ensemble des sommets possibles de départ du policier.

Démonstration. Clairement, $\cos_{\pi_0}^*$ est bornée par la probabilité de capture dans le jeu de policier et voleur puisque ce dernier peut être vu comme une relaxation de contraintes du premier, c'est-à-dire défini en laissant tomber la contrainte de visibilité du perdu. Dans le MDP \mathcal{M} , la probabilité optimale de capture est obtenue en déterminant la position initiale optimale du policier, la position maximisant sa probabilité de capture considérant la distribution initiale du voleur. Comme la distribution initiale du voleur est donnée par poc_{π_0} , le policier moyenne sa probabilité de capture sur l'ensemble des positions de départ $r \in V$ possibles du voleur, chacune d'elle apparaissant avec probabilité poc_{π_0} . \square

La proposition 4.7.1 n'est utilisable qu'au début du jeu, avant que les positions initiales soient choisies. Comme on veut appliquer une méthode de *branch and bound*, on doit pouvoir appliquer le même raisonnement, peu importe le nombre de sommets parcourus par le chercheur. Autrement dit, on cherche à obtenir une borne utilisable pour chaque chemin π_t pour toute

valeur de $t \in \{0, \dots, T\}$. La prochaine proposition étend adéquatement la proposition précédente.

Proposition 4.7.2. *Soit $\pi_t = v_0, v_1, v_2, \dots, v_t$. Alors,*

$$\cos_{\pi_t}^* \leq \cos_{\pi_t} + \max_{c' \in N[v_t]} \sum_{r' \in V} \text{poc}_t(r') u^*(r', c', t + 1). \quad (4.19)$$

On interprète ici $N[v_0]$ comme la position initiale du policier.

Démonstration. Tel que dans la démonstration précédente, $\cos_{\pi_t}^*$ est borné supérieurement par le gain optimal obtenu en jouant en premier le jeu de l'OSP sur le chemin π_t puis, si le perdu n'a pas encore été détecté, en remplaçant l'OSP par le jeu du policier et voleur saoul. Ceci est réalisé simplement en rendant le perdu visible au policier, autrement dit en jouant le jeu du policier et voleur avec le perdu dans le rôle du voleur et le chercheur dans celui du policier. Lorsque ce jeu commence, au tour $t + 1$, le policier doit sélectionner sa position de départ de manière à maximiser sa probabilité de capture; comme la position courante du voleur n'est pas connue, le policier doit moyenner la probabilité de capture du voleur sur ses positions possibles $r' \in V$, chacune étant choisie avec probabilité $\text{poc}_{\pi_t}(r')$. Ainsi, si le policier choisit l'action $c' \in N[v_t]$, sa probabilité de capture sera donnée par l'équation

$$\sum_{r' \in V} \text{poc}_{\pi_t}(r') u^*(r', c', t + 1).$$

Le résultat suit du fait que le policier cherche l'action maximisant cette équation et du fait qu'on a considéré deux événements disjoints : soit le perdu a déjà été détecté (avec probabilité \cos_{π_t}), soit il est en vie auquel cas on peut borner la probabilité qu'il soit éventuellement capturé. \square

Remarque 4.7.3. *Un des aspects importants de la borne COPWIN est sa capacité à être calculée complètement avant l'exécution du programme, c'est-à-dire hors-ligne. Autrement dit, on peut calculer chaque valeur $u^*(r, c, t)$ pour toute position de départ possible du policier et du voleur $r, c \in V$ ainsi que pour tout incrément de temps $t \in \{1, \dots, T + 1\}$ et emmagasiner les valeurs dans un tableau. Bref, lors de la résolution de l'OSP, le calcul de $u^*(r, c, t)$ se fait à coût constant. Ainsi, la borne de l'équation (4.19) ne dépend que des sommets voisins de la position courante du chercheur et donc la borne nécessite un coût de $O((\Delta + 1)n)$ sur un graphe à n sommets. En effet, tout sommet de G a au plus Δ voisins, puis pour chacun d'entre eux on doit calculer la probabilité que le voleur se retrouve à un sommet donné ce qui nécessite n opérations. Comme la complexité de la borne DMEAN est $O((T - k + 1)(n + m))$, la borne COPWIN est plus rapide sur la plupart des incréments de temps et la plupart des graphes.*

Le jeu de policier et voleur saoul fut présenté par Pralat et Kehagias en 2012 [25]. Dans cet article, les auteurs présentent aussi une description de ce jeu sous la forme d'un MDP,

cependant la formulation est différente de celle présentée ici et est orientée vers l’objectif de calculer le nombre moyen de tours joués avant la capture du voleur. Komarov et Winkler [27] ont aussi présenté ce jeu, mais sans le modéliser. Dans cet article, les auteurs développent une stratégie maison, sans démontrer formellement son optimalité, pour obtenir une borne universelle sur le temps de capture moyen du voleur par le policier pour n’importe quel graphe.

Pour appliquer la borne COPWIN en optimisation par contraintes (CP), il est nécessaire de filtrer la borne supérieure de la variable Z correspondant à la fonction objectif. Il est démontré plus haut que la borne COPWIN est admissible, c’est-à-dire elle ne sous-estime jamais la vraie valeur objectif. Dans la prochaine section sont présentés les résultats d’expérimentations réalisées pour mesurer l’efficacité de cette borne. Celle-ci a été implémentée sur un modèle d’optimisation par contraintes et comparée à la borne DMEAN présentée à la section 4.5.

4.8 Expérimentations et résultats

Les expérimentations ont toutes été menées sur le superordinateur Colosse de l’Université Laval et utilisent le solveur par contraintes Choco 2.1.5 [29] ainsi que la librairie Java Universal Network/Graph (JUNG) 2.0.1 [40] de théorie des graphes. Le temps CPU alloué à chaque instance était de 5 minutes ou 5 millions de retours-arrières.

Les instances ont été construites à l’aide de six différents graphes, dont l’habituelle grille carrée (un sous-graphe fini de \mathbb{Z}^2), nommée ici G^+ . Aussi présent est la grille G^+ auquel ont été ajoutées toutes les diagonales dans les carrés, nommé G^* . La grille hexagonale, similaire à G^+ , mais avec des hexagones remplaçant les carrés est aussi présente sous le nom G^H . Une dernière grille, nommée G^V , est simplement la grille G^H auquel est ajouté un sommet par hexagone relié à chaque sommet de l’hexagone. À ces graphes sont ajoutés le graphe G^L modélisant les tunnels de l’Université Laval et le graphe G^T qui est une transformation du graphe décrivant le métro de Londres (le Tube) et dans lequel les voisins d’un sommet de degré 2 sont connectés ensemble alors que le sommet lui-même était retiré. Certains de ces graphes sont présentés à la figure 4.4.

Aussi, pour varier les instances, les probabilités de détection, la matrice de transition ainsi que le nombre maximal de tours alloués ont été variés. Notamment, pour chaque instance on a choisi $p \in \{0.3, 0.6, 0.9\}$ et défini $\text{pod}(r) = p$ pour chaque sommet r . Ensuite, pour la matrice de transition, une valeur $\rho \in \{0.3, 0.6, 0.9\}$ est sélectionnée, puis la matrice est construite en plaçant une probabilité ρ de rester sur un sommet v et probabilité $1 - \frac{\rho}{\deg(v)}$ de transiter vers un voisin de v . Ainsi, le perdu a toujours une certaine probabilité de rester sur place et, s’il se déplace, choisit sa prochaine position selon une distribution uniforme. Finalement, les différentes valeurs de T ont été sélectionnées dans $\{5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25\}$. Au total ont été générées 594 instances.

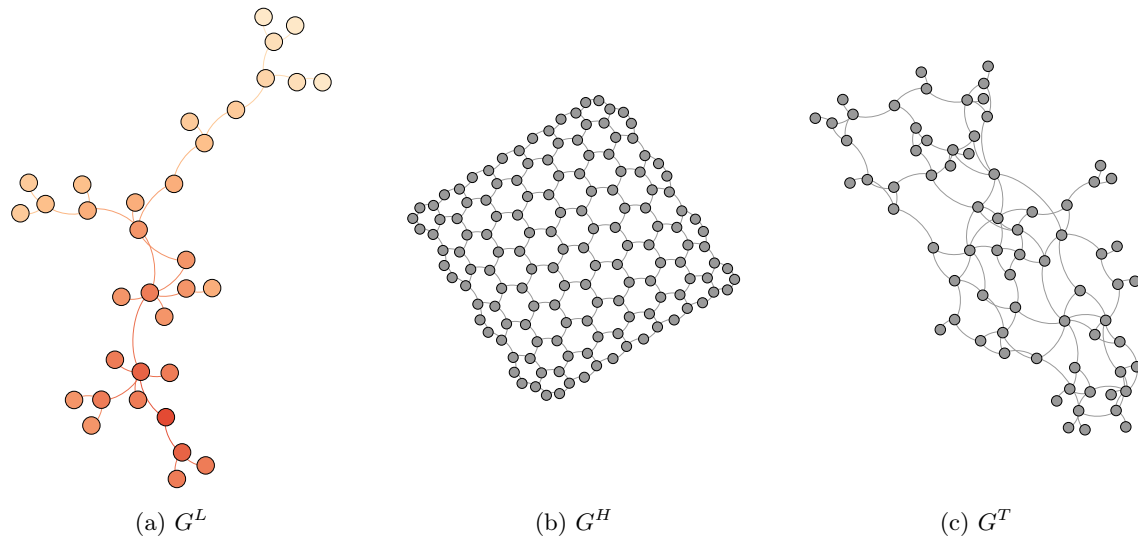


FIGURE 4.4 – Certains des graphes utilisés pour les expérimentations.

Les expérimentations ont été conduites en trois étapes. Dans les deux premières étapes, la borne COPWIN est comparée à la borne DMEAN présentée à la section 4.5. En premier lieu, les deux bornes ont été implémentées séparément comme contraintes sur la borne supérieure de la fonction objectif. Les bornes étaient ensuite évaluées sur l'ensemble des instances et les valeurs objectifs finales recueillies. Finalement, les valeurs finales calculées pour chaque borne étaient comparées sur l'ensemble des instances.

La deuxième expérimentation avait pour objectif de mesurer l'habilité de chaque borne à élaguer des branches de l'arbre de recherche. Ainsi, les deux bornes ont été implémentées côte à côte sous la forme d'une seule contrainte sur la valeur objectif. Les calculs de chacune étaient réalisés simultanément et la valeur minimale des deux bornes était utilisée pour mettre à jour la borne supérieure de la variable de la fonction objectif Z . Autrement dit, lorsque la valeur minimale des deux bornes était inférieure à la borne supérieure de la variable Z , cette borne supérieure devait être modifiée. Si le domaine de Z était vidé, alors nécessairement la valeur minimale des deux bornes était inférieure à la borne inférieure de Z , ce qui menait directement à un retour-arrière. Ainsi, au moins une des deux bornes, COPWIN ou DMEAN, devait avoir vidé le domaine de Z . Chaque fois que cela se produisait, la borne vidant le domaine voyait son score incrémenté de 1 (partant de 0) pour compter le nombre de fois que celle-ci a mené à un retour-arrière. Dans tous les cas, l'arbre de recherche est construit naturellement de telle sorte que chaque niveau corresponde à un tour de jeu et que chaque niveau soit ordonné en ordre croissant à partir de la racine, c'est-à-dire la racine a niveau 0. Ceci nous permettait de savoir exactement à quel tour de jeu un retour-arrière était déclenché par quelle borne et donc d'évaluer la capacité de chacune d'elle à élaguer des branches de l'arbre plus tôt ou plus tard dans la partie.

La troisième partie peut être vue comme une sous-expérimentation de la première. Dans celle-ci, nous avons pour objectif d'évaluer la performance de chaque borne COPWIN et DMEAN lorsque celles-ci sont accompagnées d'une heuristique efficace faisant partie de l'état de l'art dans la littérature du problème OSP. Bref, l'heuristique *Total Detection* (TD) [37] a été ajoutée au solveur lors de la première expérimentation. Cette heuristique est en fait très similaire à la borne COPWIN, cependant (TD) est une heuristique alors que la borne a été théoriquement démontrée comme étant valide. L'heuristique TD est une heuristique de choix de valeur et assigne des valeurs aux variables de chemin $PATH_t$. Même lorsque cette heuristique est ajoutée à l'expérimentation, la borne COPWIN reste compétitive vis-à-vis de la borne DMEAN.

4.8.1 Résultats de la comparaison entre COPWIN et DMEAN

À la figure 4.5 est présenté un graphique montrant les meilleures valeurs objectifs (COS) retournées sur chaque instance par le solveur implémentant tour à tour la borne COPWIN puis la borne DMEAN. COS est une estimation de la valeur optimale $\cos_{\pi_0}^*$ obtenue à la fin du processus de résolution, c'est-à-dire soit après que la valeur optimale a été atteinte, que le temps imparti soit terminé ou que le nombre de retours-arrières soit dépassé. Sur cette figure, plus un point se rapproche de la diagonale, plus la valeur correspondante est similaire pour chaque borne. Autrement, un point en dessous de la diagonale est à l'avantage de la borne COPWIN : cela démontre que la valeur calculée par cette borne est meilleure que celle calculée par la borne DMEAN. L'inverse est aussi vrai lorsqu'un point se retrouve au-dessus de la diagonale, ce qui se produit sur quelques instances. Il apparaît que la borne COPWIN mène généralement à de meilleures valeurs objectifs sur la majorité des instances. Cela est très positif pour la borne COPWIN puisqu'en plus d'être plus rapide que la borne DMEAN, tel que noté dans la remarque 4.7.3, la borne COPWIN est aussi assez précise pour diriger le processus de résolution vers de meilleures valeurs objectifs.

Remarque 4.8.1. *Il apparaît clairement que certaines valeurs sont totalement à l'opposé de la tendance générale sur le graphique de la figure 4.5 (principalement les quelques points agrégés autour d'un cône vertical d'extrémité (0.01, 0.625)). Ceci n'est pas particulièrement négatif pour la borne COPWIN, en fait une plus grande concentration de l'ensemble des points sur la diagonale serait un aspect plus négatif. En effet, pour ces points l'écart entre la valeur objectif calculée par COPWIN et celle calculée par DMEAN est tellement grand que ces valeurs sont très probablement dues à des erreurs numériques. En effet, le solveur utilisé, Choco 2.1.5, n'implémente pas de nombres réels et tous les calculs ont dû être réalisés en nombres entiers. Toutes les probabilités ont donc dû être multipliées par un grand nombre entier pour permettre différents calculs probabilistes, puis redivisées pour être renvoyées dans l'intervalle [0, 1]. Bref, il est très probable que des erreurs numériques se soient incrustées dans les calculs. C'est d'ailleurs une raison pour laquelle dans les extensions de ce projet on compte une nouvelle*

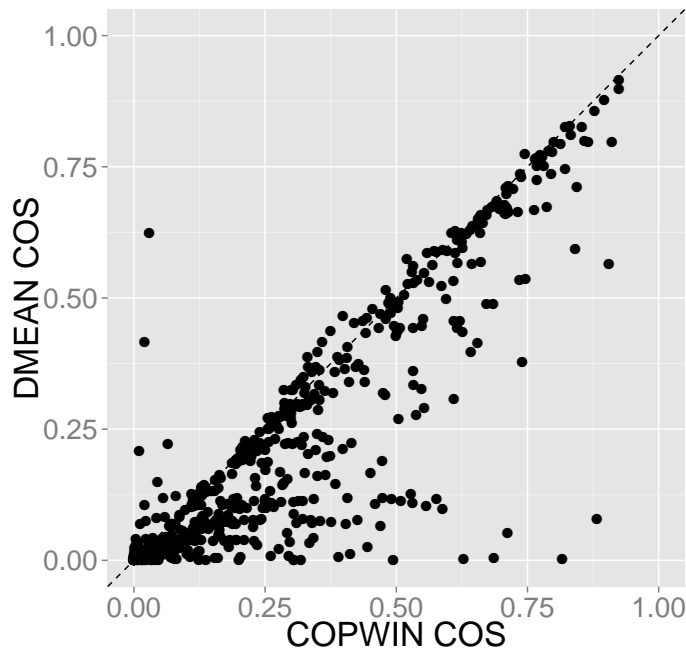


FIGURE 4.5 – Comparaison des valeurs objectifs retournées par les bornes COPWIN et DMEAN. Les 594 instances sont représentées, chaque point ● représentant la valeur objectif finale d’une seule instance.

implémentation complète de ce programme dans un solveur avec nombres réels. Dans tous les cas, pour citer un terme statistique, ces points constituent des outliers, des valeurs extrêmes qui ne suivent pas la tendance générale.

Les graphiques de la figure 4.6 montrent le nombre de fois que chaque borne a mené à une coupe de branche dans l’arbre de recherche du solveur, soit que chaque borne a vidé le domaine de la valeur objectif. Chaque graphique montre ceci pour un graphe particulier utilisé. Les différents résultats pour chaque valeur de paramètre ρ et pod ont été agrégés, seuls les graphes ont été considérés séparément. Les graphiques montrent le nombre de coupes de branche selon le temps (tour) de jeu pour mieux analyser la capacité de chaque borne à couper plus ou moins tard dans l’arbre de recherche, soit à mieux élaguer des chemins plus ou moins tard dans le jeu. Ainsi, il semble que la borne COPWIN en général mène à plus de coupes de branches que la borne DMEAN indépendamment du graphe sur lequel est joué l’OSP. Ceci étant dit, il apparaît parfois que DMEAN vide aussi souvent le domaine de Z que COPWIN tel que sur le graphe G^L , ceci n’est encore une fois pas tout à fait négatif. On doit se rappeler que la borne DMEAN est importante dans la littérature, est très performante et donc que la borne COPWIN soit compétitive envers DMEAN est une très bonne indication de sa performance. Bref, COPWIN performe très bien sur la majorité des instances et, comme on le voit en couplant ces résultats avec ceux de la première expérimentation, mène à de meilleures valeurs

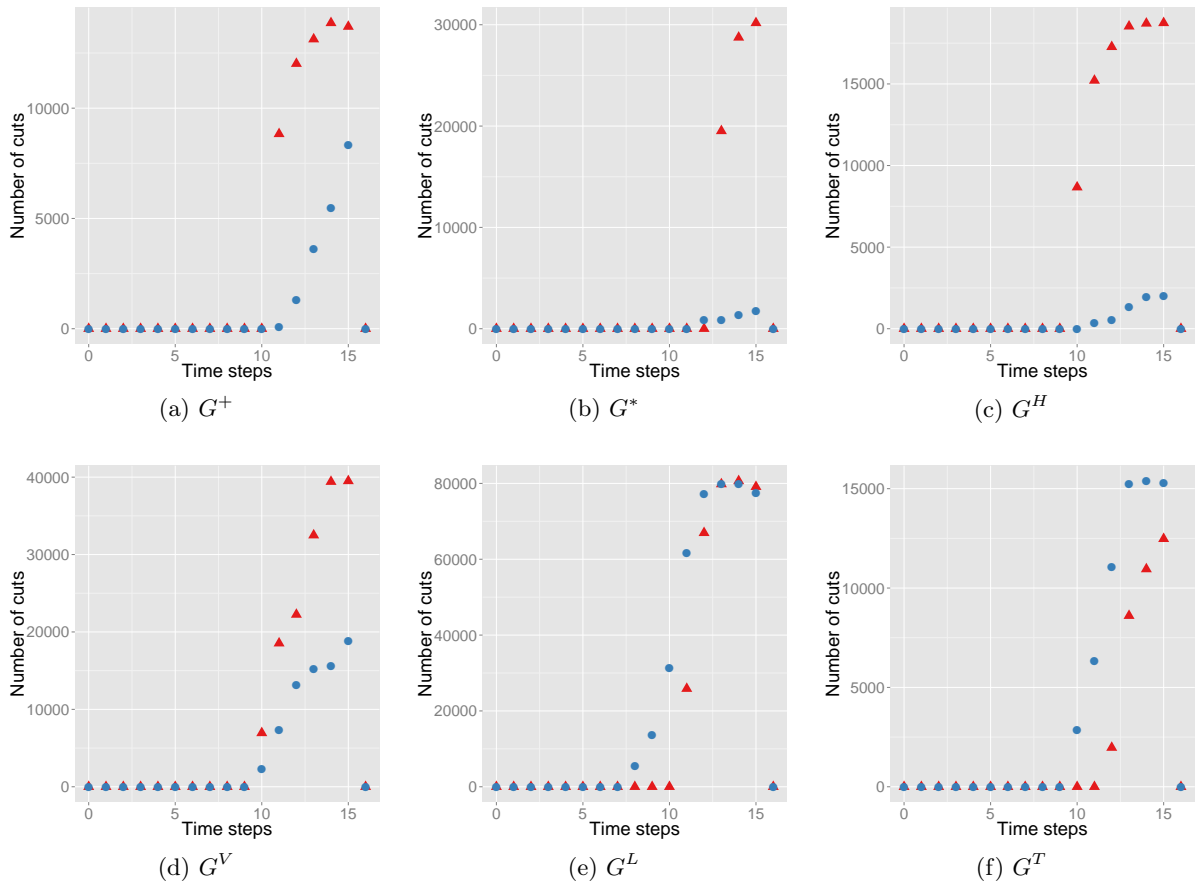


FIGURE 4.6 – Le nombre de coupes dans l’arbre de recherche réalisée par chaque borne COPWIN et DMEAN versus le tour de jeu courant avec $T = 17$. Toutes les instances ont été agrégées sur les paramètres du problème ρ et pod. Les triangles rouges \blacktriangle représentent le nombre de coupes de la borne COPWIN alors que les ronds bleus \bullet celui de la borne DMEAN.

objectifs en élaguant plus de branches que DMEAN.

La table 4.7 est ajoutée aux figures précédentes et contient l’ensemble des résultats de l’expérimentation comparant la borne COPWIN à la borne DMEAN sur tous les graphes avec $T = 17$.

Finalement, les graphiques de la figure 4.8 présentes les meilleures valeurs objectifs calculées en fonction du temps pour chaque borne COPWIN et DMEAN. Ces comparaisons sont réalisées sur trois instances représentatives : une dans laquelle COPWIN est meilleure que DMEAN, une autre dans laquelle COPWIN est pire et une dernière dans laquelle les deux bornes performant similairement. Ces graphes confirment notre intuition venant de la table 4.7 selon laquelle COPWIN permet de trouver de meilleures solutions plus rapidement que DMEAN.

4.8.2 Résultats sur l'efficacité de COPWIN

Le graphique de la figure 4.9 présente les résultats de la dernière expérimentation. Celle-ci est similaire à la première, mais au solveur a été ajoutée l'heuristique *Total Detection* (TD) [37] qui a déjà été présentée par Michael Morin et coll. en 2012 et fait partie de l'état de l'art sur le problème de l'OSP. Bien que les deux bornes guident le solveur vers de bonnes solutions, on remarque encore une fois qu'en moyenne les solutions obtenues par COPWIN sont légèrement meilleures et conservent le comportement constaté dans le graphique de la figure 4.5.

4.9 Conclusion et travaux futurs

Dans ce chapitre fut présenté un problème de la théorie de la recherche étudié sous la lunette de la programmation par contraintes. La première contribution, quoique mineure, de ce chapitre est la première implémentation de la borne DMEAN de la littérature en théorie de la recherche en optimisation par contraintes. Notre contribution principale est le développement de la borne COPWIN, une nouvelle borne sur la fonction objectif de l'OSP utilisant le concept des MDPs. Cette borne est basée sur une simple et élégante relaxation de contrainte, une relaxation en information parfaite, de ce problème de recherche vers un jeu existant de la théorie des graphes. La borne COPWIN est calculable complètement hors-ligne en temps polynomial et augmente les performances du solveur sur la majorité de nos instances. Malgré le développement de cette borne dans le monde de la programmation par contraintes, l'approche présentée dans ce chapitre peut être généralisée à d'autres méthodes de résolution de l'OSP.

Le chapitre 3 présente une généralisation des jeux de policiers et voleurs incluant le jeu du policier et voleur saoul et ses variantes. Ainsi, suivant la théorie présentée dans ce chapitre, le chapitre 3 suggère une méthode générale fondée sur la borne COPWIN pour résoudre l'ensemble des jeux de type OSP. Ainsi, Michael Morin et l'auteur de ce mémoire travaillent à la conception d'une architecture générale centrée sur le solveur Gecode⁹(en C++) pour implémenter ces idées. Il manque encore cependant une généralisation de la proposition 4.7.2 nécessitant l'encodage de la fonction objectif d'un problème OSP général, ce qui reste un léger problème ouvert.

9. Le solveur Gecode est choisi pour des raisons techniques. En effet, au moment de l'écriture de ce mémoire, il était encore problématique d'utiliser le solveur Choco 3, écrit en Java, pour faire des calculs avec des nombres réels. Gecode, cependant, est bien connu pour sa capacité à utiliser des nombres réels.

pod(r)	ρ	Objective (COS)		Time [†] (s)		Backtracks	
		DMEAN	COPWIN	DMEAN	COPWIN	DMEAN	COPWIN
<i>G^T with T = 17</i>							
0.3	0.3	0.111	0.113	298	197	3578	2791
	0.6	0.139	0.129	200	144	2045	2157
	0.9	0.362	0.354	275	208	2973	3304
0.9	0.3	0.252	0.258	131	287	1400	4613
	0.6	0.318	0.325	259	236	3082	3544
	0.9	0.736	0.736	292	185	3543	2706
<i>G⁺ with T = 17</i>							
0.3	0.3	0.011	0.014	96	37	283	118
	0.6	0.007	0.036	88	281	243	1307
	0.9	0.001	0.022	90	111	276	497
0.9	0.3	0.030	0.077	95	296	290	1394
	0.6	0.020	0.103	95	272	283	1229
	0.9	0.001	0.047	26	262	19	1174
<i>G^V with T = 17</i>							
0.3	0.3	0.035	0.079	154	296	2510	7240
	0.6	0.038	0.105	146	206	2510	4922
	0.9	0.036	0.229	158	196	2510	4898
0.9	0.3	0.099	0.215	159	240	2510	5539
	0.6	0.112	0.304	155	264	2510	6040
	0.9	0.109	0.531	161	219	2510	4996
<i>G^L with T = 17</i>							
0.3	0.3	0.344	0.318	273	145	9224	7749
	0.6	0.417	0.359	287	139	9666	7476
	0.9	0.516	0.480	297	122	9298	6462
0.9	0.3	0.638	0.651	278	119	11304	6249
	0.6	0.713	0.713	289	146	11603	7890
	0.9	0.811	0.833	249	162	9489	8727
<i>G[*] with T = 17</i>							
0.3	0.3	0.069	0.072	270	286	1264	2264
	0.6	0.107	0.117	273	183	1261	1482
	0.9	0.235	0.324	276	281	1254	2461
0.9	0.3	0.192	0.205	274	252	1264	2077
	0.6	0.304	0.333	264	219	1261	1842
	0.9	0.671	0.711	259	231	1253	1925
<i>G^H with T = 17</i>							
0.3	0.3	0.023	0.087	258	241	522	618
	0.6	0.015	0.122	255	277	519	742
	0.9	0.001	0.318	250	233	514	623
0.9	0.3	0.064	0.227	274	258	522	686
	0.6	0.043	0.342	260	286	520	719
	0.9	0.003	0.816	248	280	516	680

† Le temps requis pour calculer la dernière solution.

FIGURE 4.7 – Résultats sur différentes instances du problème OSP ; les valeurs en gras sont meilleurs.

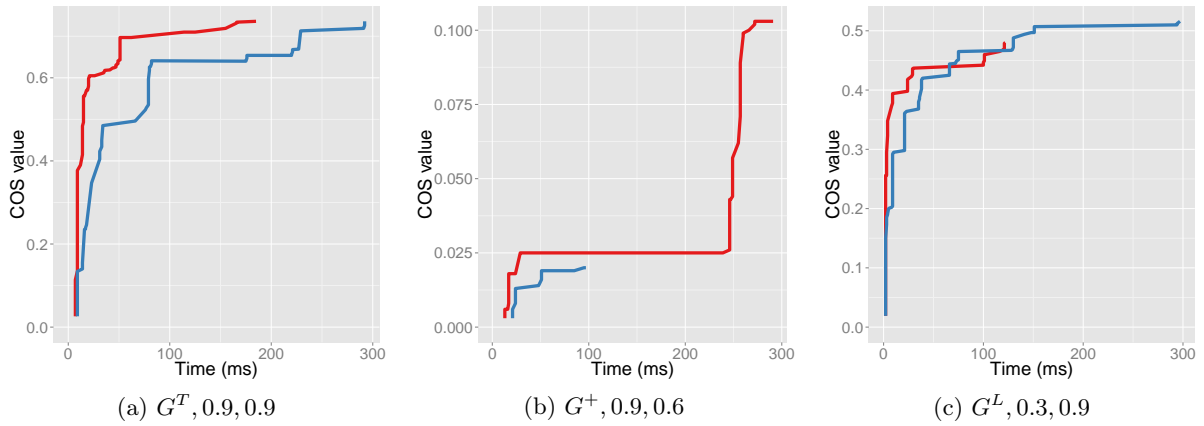


FIGURE 4.8 – La valeur objectif COS obtenue et le temps auquel elle a été obtenue pour chaque borne. Les lignes rouges — représentent la borne COPWIN alors que les lignes bleues — la borne DMEAN. Chaque graphique fait référence à un graphe particulier, la valeur pod de l’instance et sa valeur ρ . Le nombre maximal d’incrément de temps est 17.

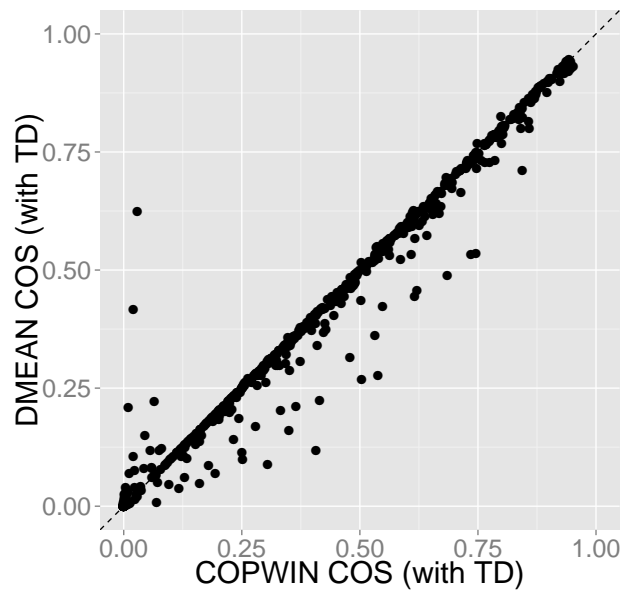


FIGURE 4.9 – Comparaison des valeurs objectifs finales obtenues avec l’heuristique TD utilisée avec les bornes COPWIN et DMEAN. Chaque point ● représente la valeur finale calculée sur une instance, l’ensemble des 594 instances est représenté.

Conclusion et ouvertures

Dans ce mémoire ont été présentés deux aspects atypiques de la théorie des jeux de policiers et voleurs. D'un premier côté, un modèle avec sa méthode de résolution a été proposé pour englober l'ensemble des variantes possibles dans une seule formulation mathématique pour, on le souhaite, faciliter l'étude et l'étendue des jeux de policiers et voleurs. D'un autre côté, il a été montré que les jeux de policiers et voleurs peuvent être utilisés de manière assez simple pour contribuer à résoudre un problème très concret qu'est le problème de l'OSP qui provient de la théorie de la recherche.

Le chapitre 3 présente un modèle qui englobe l'ensemble des jeux de policiers voleurs connus ainsi qu'une infinité d'autres. Une relation relativement simple pour le calcul de la stratégie optimale des policiers est démontrée, la récurrence w_n . Celle-ci généralise adéquatement la relation \preceq_n traditionnellement employée dans les jeux déterministes (comme dans le modèle de Bonato et MacGillivray). Par le fait même, l'idée derrière l'induction à rebours de Kuhn est remise à contribution ce qui crée un lien intéressant entre la théorie des jeux classique et celle des jeux de policiers et voleurs.

Quant à lui, le chapitre 4 présente un problème de la théorie de la recherche accompagnée d'une borne supérieure sur sa fonction objectif qui fut appliquée dans le cadre d'une méthode de *branch and bound*. Il a été montré que cette borne est valide au niveau théorique, soit qu'elle ne sous-estime jamais la valeur objectif. De plus, des résultats empiriques justifient son utilisation en pratique par la qualité de la valeur objectif calculée par cette borne ainsi que sa rapidité de calcul.

OSP et extensions

Il semble de prime abord que le seul fil conducteur entre les chapitres 3 et 4 soit le thème très général des jeux de policiers et voleurs. Ce n'est pas le cas. En creusant un peu, on découvre le fort lien qui unit ces chapitres. Car au-delà du fait qu'ils parlent tous deux de policiers et de voleurs, le fait est que l'un présente une méthode fort simple pour utiliser un jeu pour accélérer la résolution d'un problème de recherche qui est conceptuellement indépendante des règles exactes du jeu. En effet, si l'OSP avait été joué avec plusieurs chercheurs, la borne COPWIN, quoique rendue inadmissible dans sa forme actuelle, pourrait être aisément réécrite

en suivant les règles du jeu multi-policiers et voleur saoul. Mais, les variantes du problème de l'OSP sont nécessairement au moins aussi nombreuses que les variantes des jeux de policiers et voleurs. Dès lors, la généralisation du chapitre 3 prend tout son sens : la relation w_n est définie pour n'importe quel jeu de policiers et voleurs et donc la borne COPWIN peut être écrite pour n'importe quelle variante du problème OSP. Oui, il resterait cependant à démontrer pour chacune de ces variantes que la borne est effectivement admissible tel qu'il a été réalisé au chapitre 4 et c'est d'ailleurs là une des avenues de recherche qui étendrait les résultats de ce mémoire. Toutefois, on voit maintenant le travail à entreprendre. Bref, la première ouverture de ce mémoire s'énonce sous l'objectif suivant :

Démontrer qu'une extension de la borne COPWIN définie à partir de la relation w_n du chapitre 3 est admissible pour tout problème OSP respectant les mêmes règles que la relation, hormis celles concernant la visibilité des voleurs.

Toujours dans l'idée d'exploiter la symbiose implicite entre les chapitres 3 et 4, une autre direction pourrait être suivie dans des travaux futurs, celle de la généralisation du modèle 3.2.1 aux jeux à information imparfaite. La théorie des jeux suggère déjà des pistes intéressantes pour amorcer une telle étude, soit d'empêcher les joueurs (ou un seul des deux) de connaître l'état exact. Plutôt, les joueurs n'auraient accès qu'à l'information selon laquelle l'état courant est élément d'un ensemble qui, lui, serait connu des joueurs. L'ensemble des états dans lesquels peuvent se retrouver les joueurs une fois les actions choisies serait partitionné de telle sorte que chaque partition contienne plus d'un élément. Comme les joueurs ne pourraient distinguer l'élément d'une partition qui est l'état réel, ils devraient en tout temps choisir une action qui soit optimale, peu importe l'état réel. Une fois ce nouveau modèle défini, une généralisation de la relation w_n , disons w_n^i pourrait être conçue et, encore une fois, les idées derrière la borne COPWIN pourraient être exploitées. En effet, avec une relation w_n^i pourrait être démontré un résultat de la forme $w_n^i \leq w_n$ signifiant que la probabilité de capture des voleurs est toujours plus grande dans un jeu à information parfaite quand dans un jeu à information imparfaite. La relation w_n^i pourrait être utilisée pour représenter la fonction objectif d'un problème de recherche comme l'OSP. Cette idée a été exploitée dans la conception de la borne COPWIN : dans ce cas-là les deux extrêmes sur la quantité d'information ont été utilisés, un jeu à information nulle a été comparé à un jeu à information parfaite. On peut imaginer vouloir étendre ces idées en comparant un jeu à information nulle à un jeu à information partielle, dans lequel les policiers auraient certaines informations sur la position des voleurs sans connaître leurs positions exactes. La question suivante résume ceci :

Est-il possible de généraliser la relation w_n des jeux à information parfaite à une relation w_n^i pour des jeux à information imparfaite et, si oui, comment exploiter cette relation pour obtenir une borne de la forme $w_n^i \leq w_n$?

Le coût de l'aléatoire

Le chapitre 4 présente un jeu dans lequel un seul chercheur est envoyé à la recherche d'un perdu. En pratique, on peut imaginer que ceci serait insuffisant puisque le temps requis par un seul chercheur pour atteindre son objectif serait trop élevé. On remarque aussi qu'un aspect négligé dans ce chapitre est la structure du graphe sur lequel le jeu se déroule. En effet, la structure d'un graphe sur lequel un seul policier peut capturer un voleur n'est pas la même que celle d'un graphe sur lequel au moins deux policiers sont nécessaires. Pour prendre en compte le déplacement aléatoire du voleur dans l'étude du nombre de policiers nécessaire à la capture de celui-ci, Pralat et Kehagias [24] ont étudié le paramètre dit *coût de l'aléatoire* (de l'anglais *cost of randomness*) défini comme le rapport entre le temps de capture du voleur jouant de manière optimale et du temps moyen de capture du voleur saoul, soit $F(G) := \frac{T_G}{\mathbb{E}[RT_G]}$, où T_G est le temps de capture du voleur optimal et RT_G celui du voleur saoul. Dans ce rapport, on suppose que le nombre de policiers en patrouille est $c(G)$, le *cop number*, autrement T_G serait non borné. On pourrait cependant s'inspirer de ce paramètre pour étudier un rapport analogue $F_{c,c'}(G) = \frac{\mathbb{E}[RT_G^c]}{\mathbb{E}[RT_G^{c'}]}$, où RT_G^d est le temps de capture moyen du voleur saoul lorsque d policiers sont en patrouille. Ainsi, si $c = c(G)$ alors plus $F_{c(G),c'}$ diminue rapidement lorsque c' se rapproche de 1, plus il est crucial d'envoyer un nombre important de policiers en patrouille. Ce paramètre est naturellement étendu au cas du voleur invisible tel que dans l'article de Pralat et Kehagias [24] pour simuler le jeu de l'OSP. Bien que ce nouveau paramètre, disons $F_{c,c'}^i(G)$ soit nettement plus difficile à étudier que le précédent, il révèle beaucoup plus d'information quant à la structure du jeu. Encore une fois, si $F_{c(G),c'}^i(G)$ est relativement stable lorsque c' tend vers 1, alors il n'est pas trop coûteux de n'envoyer qu'un seul chercheur à la recherche du perdu, et inversement. Ce paramètre pourrait donc être étudié pour déterminer le nombre de chercheurs à envoyer en patrouille dans le problème de l'OSP avec plusieurs chercheurs. De manière plus générale, des extensions de ces paramètres pourraient être définies et étudiées directement sur le jeu 3.2.1 ainsi qu'une, possible, extension de celui-ci au cas à information imparfaite.

Retour sur les jeux de policiers et voleurs

Dans ce mémoire, on n'a qu'effleuré un des aspects importants des jeux de policiers et voleurs. Les jeux de policiers et voleurs sont souvent étudiés pour les structures de graphe intéressantes qu'ils décrivent. En effet, dans les jeux déterministes, on peut parler, par exemple, de graphes *policier-gagnants*, *k-policiers-gagnants*, etc. puisque la possibilité qu'un ou k policiers gagnent contre les voleurs sur un graphe ne dépend que du graphe lui-même. Chacun de ces graphes doit avoir une structure particulière qui va généralement changer avec les règles. Ainsi, un graphe sur lequel un policier est assuré de gagner n'a pas la même structure qu'un graphe sur laquelle un policier gagne contre un voleur rapide. Les jeux de policiers et voleurs stochastiques peuvent aussi être utilisés pour caractériser des structures de graphe en considérant la condition $c(p, n)$ -

gagnant. Par exemple, on peut définir la classe des graphes sur lesquels un policier jouant contre un voleur saoul sont $c(\frac{1}{2}, 10)$ -gagnant, c'est-à-dire les graphes sur lesquels un policier peut capturer un voleur saoul avec probabilité au moins $\frac{1}{2}$ en au plus 10 tours. Ces graphes doivent donc avoir une structure permettant à une marche aléatoire d'être interrompue assez rapidement avec une certaine probabilité. Dans le cas de jeux avec voleurs saouls, on s'intéresse donc aux structures de graphes permettant le contrôle de marches aléatoires en temps fini. Évidemment, tout ceci peut directement se faire à l'aide du modèle présenté au chapitre 3. Comme un jeu décrit selon la définition 3.2.1 contient déjà beaucoup d'information, on pose formellement la question suivante.

Peut-on caractériser l'ensemble des jeux \mathcal{G} décrits par la définition 3.2.1 qui soient $c(p, n)$ -gagnants ?

D'un autre côté, comme on travaille généralement avec des variables aléatoires, certaines quantités aléatoires deviennent dignes d'intérêt. On peut écrire le nombre $\mathbb{E}[T_{\text{cop,rob}}]$ comme l'espérance du temps de capture des voleurs par les policiers lorsque les deux joueurs jouent de manière optimale. Les premiers articles sur les jeux de policier et voleur [25, 28] saoul visaient en fait à borner cette quantité sur certaines classes de graphes. On peut donc se demander s'il est possible de borner cette même valeur, mais pour de plus grandes classes de jeux à l'aide du modèle 3.2.1.

Enfin, ce mémoire se clôt avec beaucoup d'avenues de recherches qui restent à explorer. La théorie développée dans les derniers chapitres permet de poser plus de questions que celles auxquelles nous avons répondu, ce qui semble un point très positif. En effet, cela montre que les idées développées dans ce texte sont prometteuses et offrent de belles perspectives de recherche futures.

Annexe A

Annexe

A.1 Le jeu de policiers et voleurs décrit comme un jeu stochastique simple

La définition suivante présente le jeu de policiers et voleurs général de la définition (3.2.1) sous la forme d'un jeu stochastique simple.

Définition A.1.1. *Un jeu de poursuite discret, ou jeu de policiers et voleurs, \mathcal{G} se décrit par le tuple*

$$\mathcal{G} = ((c, r), V, (V_{\text{cop}}, V_{\text{rob}}), \Delta, W).$$

\mathcal{G} est un jeu stochastique à deux joueurs avec accessibilité simple. L'objectif des policiers est d'entrer dans un sommet de W .

Soit S_{cop} et S_{rob} les positions possibles des policiers et des voleurs. On pose $D = (V, E)$ comme un graphe dirigé avec ensemble de sommets $V = V_{\text{cop}} \dot{\cup} V_{\text{rob}} \dot{\cup} V_s$. Les ensembles V_{cop} et V_{rob} sont dits contrôlés par les policiers et les voleurs et sont définis par

$$\begin{aligned} V_{\text{cop}} &\subseteq S_{\text{cop}} \times S_{\text{rob}} \times \{c\} \\ V_{\text{rob}} &\subseteq S_{\text{cop}} \times S_{\text{rob}} \times \{r\}. \end{aligned}$$

L'ensemble d'arêtes E est défini implicitement plus loin à l'aide de la relation de transition Δ . On pose encore $F \subset S_{\text{cop}} \times S_{\text{rob}}$ comme l'ensemble des états de jeu finaux. Soit $s \in S_{\text{cop}} \times S_{\text{rob}}$ un état contrôlé, alors les ensembles $A_{\text{cop}}(s)$ et $A_{\text{rob}}(s)$ définissent les ensembles d'actions jouables des policiers et des voleurs. L'ensemble des sommets V_s , dits stochastiques, est donné par

$$V_s := S_{\text{cop}} \times S_{\text{rob}} \times (A_{\text{cop}}(s) \dot{\cup} A_{\text{rob}}(s)) \quad \forall s \in S_{\text{cop}} \times S_{\text{rob}}. \quad (\text{A.1})$$

Pour un sommet $v \in V_{\text{cop}} \cup V_{\text{rob}}$, on écrit $v_{\text{cop}}, v_{\text{rob}}, v_j$ pour ses composantes $S_{\text{cop}}, S_{\text{rob}}, \{c, r\}$. Similairement, pour un sommet stochastique, on écrit $v_{\text{cop}}, v_{\text{rob}}, v_a$ pour dénoter chaque composante individuelle.

On définit maintenant la relation de transition

$$\Delta \subseteq V \times (\{\perp\} \cup [0, 1]) \times V,$$

dans laquelle le symbole \perp signifie que le sommet de départ est contrôlé, mais que celui d'arrivée est stochastique ou contrôlé. Cette relation satisfait :

$$(v, \perp, w) \in \Delta \iff (v_j = w_j) \text{ et } \begin{cases} v_j = c, w_a \in A_{\text{cop}}((w_{\text{cop}}, w_{\text{rob}})) \text{ ou} \\ v_j = r, w_a \in A_{\text{rob}}((w_{\text{cop}}, w_{\text{rob}})) \text{ ou} \\ v \in V_{\text{cop}}, (w_{\text{cop}}, w_{\text{rob}}) \in F. \end{cases}$$

Cette relation doit aussi être définie lorsque le sommet de départ est stochastique, auquel cas la seconde composante de Δ est un nombre $p \in [0, 1]$ qu'on considère comme une probabilité de transition. Ainsi, lorsque $v \in V_s$, il existe un unique nombre $p \in [0, 1]$ et un sommet $w \in V_{\text{cop}} \dot{\cup} V_{\text{rob}}$ tel que $(v, p, w) \in \Delta$. On pose :

$$p > 0 \iff \begin{cases} v = (v_{\text{cop}}, v_{\text{rob}}, v_a) \text{ et } w \in S_{\text{cop}} \times \{v_{\text{rob}}\} \times \{c\}, \text{ ou} \\ v = (v_{\text{cop}}, v_{\text{rob}}, v_a) \text{ et } w \in \{v_{\text{cop}}\} \times S_{\text{rob}} \times \{r\}. \end{cases}$$

L'ensemble W contient les sommets de jeu terminaux, soit les sommets sur lesquels les voleurs sont capturés et la partie se termine. Cet ensemble est :

$$W := \{v \in V_{\text{rob}} : (v_{\text{cop}}, v_{\text{rob}}) \in F\}.$$

Remarque A.1.2. Quelques remarques sont nécessaires pour bien comprendre ce modèle.

- L'utilisation des jeux stochastiques n'enlève rien au fait que la définition (2.2.2) doit utiliser des ensembles d'états abstraits S puisqu'à ce moment-ci la structure du jeu n'est pas encore précisée. Le jeu stochastique décrit donc une structure venant s'ajouter au jeu abstrait (3.2.1).
- Les sommets stochastiques, définis à l'équation (A.1), doivent inclure les ensembles d'action de chaque joueur ainsi qu'un état de jeu puisqu'à cette étape d'une partie toute l'information doit être préservée. Les joueurs, dans le jeu stochastique, et plus particulièrement dans le jeu (2.2.2), ne font que choisir des sommets stochastiques adjacents. Pour simuler les choix d'actions abstraites des joueurs, comme dans le jeu (3.2.1), les sommets stochastiques doivent donc contenir toutes ces informations. La probabilité de transition d'un sommet stochastique, p , peut donc être perçue comme les probabilités $T_{\text{cop}}(s, a, s')$ ou $T_{\text{rob}}(s, a, s')$ définies dans le jeu abstrait.

Ainsi, les théorèmes qui suivent peuvent être déduits de la définition (2.2.2). On note σ une stratégie quelconque et $\bar{\sigma}$ une paire de stratégies pour les policiers et les voleurs. Le nombre $\mathbb{P}[W \mid \bar{\sigma}, v]$ est la probabilité qu'un sommet v mène à un sommet gagnant de W lorsque la paire de stratégies $\bar{\sigma}$ est utilisée. Cette probabilité est bien définie du fait qu'une paire de stratégies détermine entièrement le déroulement d'une partie de \mathcal{G} . La *valeur* d'un jeu est décrite dans la définition suivante. Celle-ci est définie pour donner la probabilité optimale que les policiers gagnent, peu importe la stratégie utilisée par les deux joueurs.

Définition A.1.3. *Soit \mathcal{G} le jeu de policiers-voleurs défini précédemment. Étant donné une stratégie τ et un sommet v , la valeur de la stratégie τ pour les policiers à partir du sommet v est la probabilité minimale de victoire que peut garantir cette stratégie, soit :*

$$val_{\text{cop}}^{\tau}(v) := \inf_{\sigma} \mathbb{P}[W \mid \sigma, \tau, v].$$

La valeur d'un jeu \mathcal{G} à partir d'un sommet v est donc la probabilité maximale, parmi l'ensemble des stratégies disponibles, que les policiers capturent les voleurs. Cette valeur est donnée par :

$$val_{\text{cop}}^{\mathcal{G}}(v) := \sup_{\tau} val_{\text{cop}}^{\tau}(v) = \sup_{\tau} \inf_{\sigma} \mathbb{P}[W \mid \tau, \sigma, v].$$

Comme le jeu se déroule à deux joueurs, on a l'égalité suivante :

$$val_{\text{rob}}^{\mathcal{G}}(v) = 1 - val_{\text{cop}}^{\mathcal{G}}(v).$$

Puisque la valeur d'un jeu pour les voleurs est entièrement déterminée par la valeur des policiers, on peut laisser tomber l'indice cop et simplement écrire $val^{\mathcal{G}} := val_{\text{cop}}^{\mathcal{G}}$. Ainsi, l'existence¹ de la valeur $val^{\mathcal{G}}$ implique automatiquement l'existence de la valeur de jeu pour les voleurs.

Étant donné un sommet $v_0 \in V$, on dit d'une stratégie τ qu'elle est presque sûrement gagnante si $val^{\tau}(v_0) = 1$. Le terme presque sûrement est utilisé en probabilité lorsque la probabilité qu'une suite infinie de variables aléatoires prenne une certaine valeur tend vers 1. On utilise donc ce terme ici puisqu'une partie de \mathcal{G} se décrit avec une suite infinie de sommets. Ainsi, sous la stratégie τ , la probabilité qu'un joueur gagne tend vers 1 lorsque le nombre de sommets parcourus tend vers l'infini, il est donc possible que la probabilité que ce même joueur ait gagné après n'importe quel temps fini soit strictement inférieure à 1.

Autrement, encore à partir d'un sommet v_0 , on dit d'une stratégie τ qu'elle est ϵ -optimale pour $\epsilon > 0$ si $val^{\tau}(v_0) \geq val^{\mathcal{G}} - \epsilon$. Une stratégie optimale est une stratégie qui atteint la valeur du jeu, soit $val^{\tau}(v_0) = val^{\mathcal{G}}(v_0)$. Une stratégie est globalement (ϵ -)optimale si elle est (ϵ -)optimale pour tout sommet de départ v_0 .

1. On doit en effet démontrer l'existence de la valeur d'un jeu. On a essentiellement défini un programme d'optimisation et il est inutile de chercher à calculer une valeur optimale lorsque celle-ci n'existe pas. D'autant plus qu'il existe beaucoup d'exemples de jeux en théorie des jeux pour lesquels les valeurs n'existent pas, auquel cas on ne peut donc pas parler de stratégie optimale.

Enfin, il reste à définir le concept de jeu résiduel. Supposons qu'une partie soit déjà commencée et qu'on veuille décrire la meilleure stratégie à jouer dans le restant du jeu. Alors, comme une partie se décrit avec des sommets, le jeu doit être actuellement sur un sommet v . On appelle donc $\mathcal{G}[v]$ le jeu résiduel débutant en v , soit le jeu qui reste à jouer lorsque le sommet courant est v . Une stratégie τ_v peut être optimale de manière résiduelle en étant optimale dans le jeu résiduel $\mathcal{G}[v]$.

Ceci dit, on peut enfin utiliser quelques théorèmes de la théorie des jeux stochastiques pour les appliquer dans les jeux de policiers-voleurs. Tous ces théorèmes sont présentés sans démonstrations, mais plus de plus amples informations peuvent être trouvées dans le travail de Ummels [56].

Théorème A.1.4. *Soit \mathcal{G} le jeu décrit à la définition A.1.1. Alors, on a les résultats suivants.*

1. *La valeur $val^{\mathcal{G}}$ existe et pour tout $\epsilon > 0$, chaque joueur possède une stratégie ϵ -optimale déterministe [32, 35, 56]. Autrement dit, les policiers ont une stratégie déterministe leur garantissant une victoire avec probabilité au moins $val^{\mathcal{G}} - \epsilon$ alors les voleurs peuvent survivre avec probabilité au moins $1 - val^{\mathcal{G}} + \epsilon$ avec une stratégie déterministe, et ce pour tout $\epsilon > 0$.*
2. *Pour tout jeu résiduel de \mathcal{G} , il existe des stratégies résiduelles optimales déterministes [19, 56].*
3. *Les deux joueurs ont des stratégies optimales globales déterministes et sans-mémoire [56]. Les joueurs ont donc des stratégies qui atteignent les valeurs $val^{\mathcal{G}}$ et $1 - val^{\mathcal{G}}$ en plus de leurs stratégies ϵ -optimales. Qui plus est, celles-ci sont sans mémoire, ce qui est très important au niveau algorithmique.*

Enfin, lorsque le jeu est entièrement déterministe, c'est-à-dire les probabilités de transition des sommets stochastiques sont Diracs, on a aussi le résultat qui suit.

Théorème A.1.5. *Dans n'importe quel jeu résiduel de \mathcal{G} , soit les policiers ont une stratégie gagnante déterministe ou les voleurs en ont une [34, 56].*

Ce dernier théorème est essentiellement une réécriture d'un résultat assez connu dans les jeux combinatoires [15] qu'a inspiré le résultat de Zermelo. Celui-ci avait en effet montré [55] que dans le jeu d'Échecs, soit un joueur possède une stratégie gagnante déterministe, soit les deux joueurs sont assurés de faire une partie nulle.

Bibliographie

- [1] AIGNER, M. et FROMME, M. (1984). A game of cops and robbers. *Discrete Applied Mathematics*, 8(1):1–12.
- [2] ALSPACH, B., CLARKE, N., HAHN, G., NOWAKOWSKI, R. J., VALTCHEV, P. et YANG, B. (2002). Searching networks. <http://www.mprime.ca/fr/node/153/searching-networks>. (Consulté le 28 mai 2016).
- [3] BARTO, A. G. et SUTTON, R. S. (1998). *Reinforcement learning : An introduction*. MIT press.
- [4] BELDICEANU, N. et MASSEY, S. (2014). Global constraint catalog. <http://sofdem.github.io/gccat/>. (Consulté le 28 mai 2016).
- [5] BOLLOBÁS, B. (1998). *Modern graph theory*. Graduate Texts in Mathematics (Book 184). Springer.
- [6] BONATO, A. et MACGILLIVRAY, G. (2016). Characterizations and algorithms for generalized cops and robbers games. http://www.math.ryerson.ca/~abonato/papers/GenCandR_BG_May15.pdf. (Consulté le 28 mai 2016).
- [7] BONATO, A., MITSCHÉ, D., PÉREZ-GIMÉNEZ, X. et PRALAT, P. (2015). A probabilistic version of the game of Zombies and Survivors on graphs. *Theoretical Computer Science*, pages 1–18. (Sous presse).
- [8] BONATO, A. et NOWAKOWSKI, R. J. (2011). *The Game of Cops and Robbers on Graphs*. American Mathematical Society.
- [9] BONDY, A. et MURTY, U. (2008). *Graph Theory*. Graduate Texts in Mathematics. Springer-Verlag London.
- [10] CHALOPIN, J., CHEPOI, V., NISSE, N. et VAXÈS, Y. (2011). Cop and Robber Games When the Robber Can Hide and Ride. *SIAM Journal on Discrete Mathematics*, 25(1):333–359.
- [11] CHEN, H. et ZHANG, F. (2008). The expected hitting times for finite Markov chains. *Linear Algebra and its Applications*, 428(11-12):2730–2749.

- [12] CLARKE, N. E. et MACGILLIVRAY, G. (2012). Characterizations of k-copwin graphs. *Discrete Mathematics*, 312(8):1421–1425.
- [13] COGILL, R. et PENG, C. (2010). A spanning tree method for bounding hitting times of random walks on graphs. *SIAM Journal on Discrete Mathematics*, 24(3):808–820.
- [14] COLSON, B., MARCOTTE, P. et SAVARD, G. (2007). An overview of bilevel optimization. *Annals of operations research*, 153(1):235–256.
- [15] CONWAY, J. H. (1976). *On numbers and games*. Academic Press.
- [16] DIESTEL, R. (2010). *Graph Theory*. Graduate Texts in Mathematics. Springer Berlin Heidelberg.
- [17] FOMIN, F. V. et THILIKOS, D. M. (2008). An annotated bibliography on guaranteed graph searching. *Theoretical Computer Science*, 399(3):236–245.
- [18] FREDMAN, M. L. et TARJAN, R. E. (1987). Fibonacci heaps and their uses in improved network optimization algorithms. *J. ACM*, 34(3):596–615.
- [19] GIMBERT, H. et HORN, F. (2010). Solving simple stochastic tail games. *In Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 847–862. Society for Industrial and Applied Mathematics.
- [20] GOLDSTEIN, A. S. et REINGOLD, E. M. (1995). The complexity of pursuit on a graph. *Theoretical Computer Science*, 143(1):93–112.
- [21] GOLOVACH, P. (1989a). A topological invariant in pursuit problems. *Differential Equations*, 25(6):657–661.
- [22] GOLOVACH, P. (1989b). Equivalence of two formalizations of the search problem in a graph. *Vestnik leningradskogo universiteta seriya matematika mekhanika astronomiya*, (1):10–14.
- [23] ISAACS, R. (1999). *Differential games : a mathematical theory with applications to warfare and pursuit, control and optimization*. Dover Books on Mathematics. Dover Publications.
- [24] KEHAGIAS, A., MITSCHKE, D. et PRALAT, P. (2013). Cops and Invisible Robbers : The Cost of Drunkenness. *Theoretical Computer Science*, 481:100–120.
- [25] KEHAGIAS, A. et PRALAT, P. (2012). Some remarks on cops and drunk robbers. *In Theoretical Computer Science*, volume 463, pages 133–147.
- [26] KINNERSLEY, W. B. (2015). Cops and Robbers is EXPTIME-complete. *Journal of Combinatorial Theory, Series B*, 111:201–220.

- [27] KOMAROV, N. et WINKLER, P. (2013). Cops vs. Gambler. <http://arxiv.org/abs/1308.4715>. (Consulté le 28 mai 2016).
- [28] KOMAROV, N. et WINKLER, P. (2014). Capturing the Drunk Robber on a Graph. *The Electronic Journal of Combinatorics*, 21(3):14.
- [29] LABURTHE, F. et JUSSIEN, N. (2012). Choco solver documentation. Rapport technique, École des Mines de Nantes. http://choco-solver.org/user_guide/. (Consulté le 28 mai 2016).
- [30] LAU, H., HUANG, S. et DISSANAYAKE, G. (2008). Discounted MEAN bound for the optimal searcher path problem with non-uniform travel times. *European Journal of Operational Research*, 190(2):383–397.
- [31] LIGGETT, T. M. et LIPPMAN, S. A. (1969). Stochastic games with perfect information and time average payoff. *SIAM Review*, 11(4):604–607.
- [32] MAITRA, A. et SUDDERTH, W. (1998). Finitely additive stochastic games with borel measurable payoffs. *International Journal of Game Theory*, 27(2):257–267.
- [33] MARCOUX, H. (2014). Jeux de poursuite policier-voleur sur un graphe, le cas du voleur rapide. Mémoire de maîtrise, Université Laval.
- [34] MARTIN, D. A. (1975). Borel determinacy. *Annals of Mathematics*, pages 363–371.
- [35] MARTIN, D. A. (1998). The determinacy of blackwell games. *The Journal of Symbolic Logic*, 63(04):1565–1581.
- [36] MARTINS, G. H. (1993). *A new branch-and-bound procedure for computing optimal search paths*. Thèse de doctorat, Naval Postgraduate school.
- [37] MORIN, M., PAPILLON, A.-P., ABI-ZEID, I., LAVIOLETTE, F. et QUIMPER, C.-G. (2012). Constraint Programming for Path Planning with Uncertainty : Solving the Optimal Search Path problem. *In Principles and Practice of Constraint Programming*, pages 988–1003. Springer.
- [38] NORRIS, J. R. (1998). *Markov Chains*. Statistical and Probabilistic Mathematics. Cambridge University Press.
- [39] NOWAKOWSKI, R. et WINKLER, P. (1983). Vertex-to-vertex pursuit in a graph. *Discrete Mathematics*, 43(2-3):235–239.
- [40] O’MADADHAIN, JOSHUA FISHER, D., NELSON, T., WHITE, S. et BOEY, Y.-B. (2010). The JUNG (Java Universal Network/Graph) Framework. <http://jung.sourceforge.net>. (Consulté le 28 mai 2016).

- [41] OSBORNE, M. J. et RUBINSTEIN, A. (1995). *A course in game theory*, volume 29. MIT Press, Cambridge, Mass.
- [42] PALACIOS, J. L. et RENOM, J. M. (2010). Sum rules for hitting times of Markov chains. *Linear Algebra and its Applications*, 433(2):491–497.
- [43] PARSONS, T. D. (1978). Pursuit-evasion in a graph. *In Theory and applications of graphs*, pages 426–441. Springer.
- [44] PETROV, N. N. (1982). A problem of pursuit in the absence of information on the pursued. *Differentsial'nye Uravneniya*, 18(8):1345–1352.
- [45] PUTERMAN, M. L. (2014). *Markov decision processes : discrete stochastic dynamic programming*. John Wiley & Sons.
- [46] QUILLIOT, A. (1983). *Problèmes de jeux, de point fixe, de connectivité et de représentation sur des graphes, des ensembles ordonnés et des hypergraphes*. Thèse de doctorat d'état, Université de Paris VI, France.
- [47] ROBSON, J. M. (1984). N by N checkers is Exptime complete. *SIAM Journal on Computing*, 13(2):252–267.
- [48] ROSS, S. (2009). *Introduction to probability models*. Academic Press, 10ème édition.
- [49] ROSS, S. (2012). *A First Course in Probability*. Pearson, 9ème édition.
- [50] SHAPLEY, L. (1953). Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095.
- [51] SHOHAM, Y. et LEYTON-BROWN, K. (2008). *Multiagent systems : algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.
- [52] SIMARD, F., MORIN, M., QUIMPER, C.-G., LAVIOLETTE, F. et DESHARNAIS, J. (2015). Bounding an Optimal Search Path with a Game of Cop and Robber on Graphs. *In Principles and Practice of Constraint Programming : 21st International Conference, CP 2015, Cork, Ireland, August 31–September 4, 2015, Proceedings*, volume 9255, pages 403–418. Springer Science Business Media.
- [53] TADELIS, S. (2013). *Game theory : an introduction*. Princeton University Press.
- [54] TRUMMEL, K. et WEISINGER, J. R. (1986). The Complexity of the Optimal Searcher Path Problem.
- [55] ULRICH, S. et WALKER, P. (2001). Zermelo and the Early History of Game Theory. *Games and Economic Behavior*, 34(1):123–137.

- [56] UMMELS, M. (2010). *Stochastic Multiplayer Games : Theory and Algorithms*. Amsterdam University Press, Amsterdam.
- [57] WASHBURN, A. R. (1998). Branch and bound methods for a search problem. *Naval Research Logistics*, 45(3):243–257.
- [58] WEGENER, I. (2005). *Complexity theory : exploring the limits of efficient algorithms*. Springer Science Business Media.
- [59] WIKIPÉDIA (2015). Stratego — Wikipédia, l’encyclopédie libre. <https://fr.wikipedia.org/wiki/Stratego>. (Consulté le 28 mai 2016).