Rapport de projet

Amadou KANE – Hamza ZOUDANI

Master 1 Informatique – 2014/2015

Mise en œuvre et analyse expérimentale d’un algorithme de résolution du problème de satisfaction de contraintes

*Encadré par M. Philippe JEGOU*

****

# Sommaire

[Sommaire 1](#_Toc420674247)

[1. Problème à satisfaction de contraintes 5](#_Toc420674248)

[2. Algorithmes de résolution 6](#_Toc420674249)

[2.1. Backtrack 6](#_Toc420674250)

[2.2. Foward Checking 7](#_Toc420674251)

[2.3. Arc consistance et filtrage 8](#_Toc420674252)

[2.3.1. *RFL* 9](#_Toc420674253)

[2.3.2. *MAC* 9](#_Toc420674254)

[3. Méthode de travail 9](#_Toc420674255)

[3.1. Gestion de projet 9](#_Toc420674256)

[3.2. Langage et outils de développement 10](#_Toc420674257)

[3.2.1. Langage de programmation 10](#_Toc420674258)

[3.2.2. Travail collaboratif 11](#_Toc420674259)

[4. Travail réalisé 11](#_Toc420674260)

[4.1. Description du projet 11](#_Toc420674261)

[4.2. Description des structures créés 11](#_Toc420674262)

[4.3. Génération d’un graphe aléatoire 12](#_Toc420674263)

[4.3.1. Procédures pour générer le graphe 12](#_Toc420674264)

[4.3.2. Modélisation du problème de 3-colorabilité avec la structure 13](#_Toc420674265)

[4.4. Implémentation des algorithmes 14](#_Toc420674266)

[4.4.1. Backtrack 14](#_Toc420674267)

[4.4.2. Foward Checking 14](#_Toc420674268)

[4.4.3. RFL 15](#_Toc420674269)

[4.5. Intégration des heuristiques 16](#_Toc420674270)

[4.6. Graphes de comparaison 17](#_Toc420674271)

[4.6.1. Foward Checking 17](#_Toc420674272)

[4.6.2. RFL 18](#_Toc420674273)

Remerciements

Nous tenons à remercier tout particulièrement et à témoigner notre reconnaissance à M. Philippe Jégou, notre tuteur de projet, pour l’aide qu’il nous a apporté durant la période du projet. Sa disponibilité, sa pédagogie et ses conseils nous ont été précieux dans nos choix de conception et d d’élaboration de nos programmes.

Introduction

Dans le cadre de notre Master 1 Informatique, les étudiants sont amenés à choisir entre un projet et un stage. Si les étudiants choisissent un projet, ils doivent en choisir un parmi ceux proposés par les enseignants de l’Université d’Aix-Marseille. Nous avons choisi le projet « Mise en œuvre et analyse expérimentale d’un algorithme de résolution du problème de satisfaction de contraintes » tuteuré par M Philippe Jégou, enseignant-chercheur d’Aix-Marseille.

# Problème à satisfaction de contraintes

Un problème à satisfaction de contraintes a trois composantes : X, D et C :

* X est un ensemble de variables, {,…,}.
* D est un ensemble de domaines, {,…,}, un pour chaque variable.
* C est un ensemble de contraintes qui spécifie les combinaisons admissibles de valeurs.

Chaque domaine est constitué d'un ensemble de valeurs admissibles {,…,} pour la variable. Chaque contrainte est constituée d'une paire de (***portée***, ***rel***), où ***portée*** est un n-uplet de variables qui participent à la contrainte et ***rel*** est une relation qui définit les valeurs que ces variables peuvent prendre. Une relation peut être représentée comme une liste explicite de tous les n-uplets de valeurs qui satisfont la contrainte, ou comme une relation abstraite sur laquelle s'appliquent deux opérations:

* Tester si un n-uplet est un membre de la relation.
* Enumérer les membres de la relation.

Par exemple, si et ont tous les deux le domaine , la contrainte qui dicte que les variables doivent avoir des valeurs différentes peut s'écrire <(,), [(A, B), (B, A)]> ou <(,),#>.

L'objectif consiste simplement à trouver un ensemble de valeurs à affecter aux variables, de sorte que toutes les contraintes soient satisfaites.

Nous nous intéressons dans le cadre de ce projet aux problèmes de satisfaction de contraintes binaires en domaines finis.

Un *CSP*[[1]](#footnote-1)binaire est un CSP dont toutes les contraintes appartenant à ont une arité à 2, c'est-à-dire que chaque contrainte a exactement 2 variables pertinentes.

Exemple d’un CSP binaire : le coloriage de graphes

Le problème de coloriage de graphe avec 3 couleurs consiste à colorier un graphe non-orienté comprenant nœuds. Chaque nœud doit être colorié avec une des trois couleurs disponibles de telle sorte que deux nœuds voisins n'aient pas la même couleur.

# Algorithmes de résolution

Les algorithmes de résolution ont pour but de vérifier que toutes les variables peuvent prendre des valeurs qui appartiennent à leur domaine, tout en respectant les différentes contraintes.

Cet algorithme devra donc générer au moins une solution qui vérifie le système. Il devra pour cela, affecter des valeurs aux variables tout en vérifiant que ces valeurs ne s'opposent pas aux contraintes.

Il existe plusieurs types d'algorithmes. Certains fonctionnent de manière systématique en testant toutes les solutions possibles jusqu'à ce qu'ils trouvent une bonne solution. D'autres, plus intuitifs, construisent la réponse à partir d'une solution non satisfaisante quelconque en la modifiant selon les contraintes.

## Backtrack

Cet algorithme trouve systématiquement une solution s'il en existe une. Pour cela, il affecte au fur et à mesure une valeur de à chaque variable correspondante dans . Il vérifie évidemment que la valeur est correcte par rapport aux contraintes.

A chaque affectation, les domaines des variables restantes diminuent. Il arrive qu'à l'affectation d'une variable, l'algorithme ne puisse trouver de valeur pour car est vide.

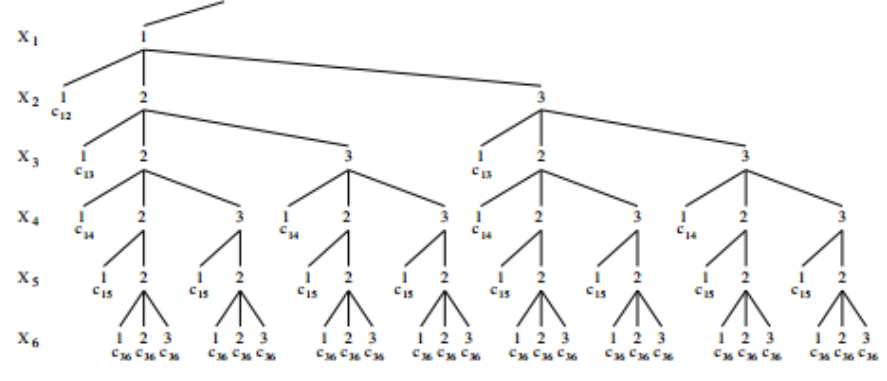
A ce stade, l'algorithme revient en arrière, à la dernière variable affectée. Il modifie la valeur de  en espérant que le domaine de la variable suivante ne sera plus nul. Si après avoir essayé toutes les valeurs du domaine , il n'a pas toujours de solutions pour , il recule et modifie la variable précédente. Si il n'existe pas de solution, il reculera jusqu’à, sinon il trouve obligatoirement la solution.

Le défaut de cet algorithme est qu'il peut tester toutes les valeurs possibles avant de trouver une solution. Pour le problème de la coloration de graphe, si on lance la procédure avec 100 sommets et 3 couleurs possibles, il existe 5,15 \* 1047 solutions. Si jamais la première bonne solution se trouve dans les dernières possibilités ou si jamais il n'y a pas de solutions,  le temps de calcul sera très long.

Exemple :

On va exécuter l’algorithme du backtrack avec les contraintes suivantes :

Le graphe suivant montre le nombre de tests qu’on aura effectué :



*Figure 1 : Graphe résumant les tests du backtrack*

## Foward Checking

Cet algorithme est assez proche du backtrack, il affecte des valeurs aux variables au fur et à mesure. La différence se trouve dans la gestion des impasses, quand l'algorithme ne trouve plus de solutions.

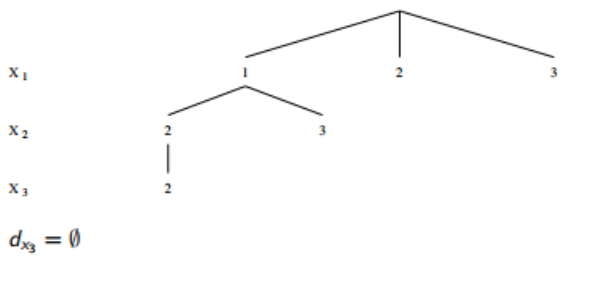
Le foward checking, avant de choisir une valeur pour une variable, vérifie que cette affectation correspond aux contraintes et que les autres variables  (i>n) pourront être affectées. Si l'affectation de ne peut être faite, l'algorithme choisit une autre valeur pour.

Si aucune solution n’est trouvée pour, on ne pourra pas faire d’affectation aux autres variables, la procédure fera un retour en arrière sur  pour changer sa valeur. Ce point reste identique au backtrack. Si le *CSP* n'a pas de solution, on recule jusqu'à la variable.

Exemple :

On va exécuter l’algorithme du foward checking avec les contraintes suivantes :

Le graphe suivant montre le nombre de tests qu’on aura effectué :



*Figure 2 : Graphe résumant les tests du foward checking*

## Arc consistance et filtrage

Un CSP (X, D, C) est consistant d’arc, si pour tout couple de variables de X, et pour toute valeur appartenant à D), il existe une valeur appartenant à D) telle que l’affectation partielle {} satisfasse toutes les contraintes binaires de C.

L'algorithme qui enlève les valeurs des domaines des variables d'un CSP jusqu'à ce qu'il soit consistant d'arc -on dit que l'algorithme filtre les domaines des variables- s'appelle *AC*[[2]](#footnote-2). Il existe différentes versions de cet algorithme: *AC1, AC2, AC3*, ..., chaque version étant plus efficace (en général) que la précédente.

### *RFL[[3]](#footnote-3)*

L’algorithme a pour objectif de résoudre des instances de *CSP* et réalise une recherche en profondeur d’abord avec les retours en arrières. A chaque étape de la recherche, une assignation de variable est effectuée suivie par un processus de filtrage qui correspond à établir la consistance d’arc.

### *MAC*[[4]](#footnote-4)

L’algorithme consiste à instancier une variable par valeur de son domaine qui est considérée comme la réduction de ce dernier à la valeur choisie ; cette réduction est donc propagée pour obtenir la consistance d'arc. Ce processus est répété lors de chaque instanciation.

Cette technique permet d'élaguer considérablement l'espace de recherche et semble (en l'état actuel des connaissances) présenter le meilleur compromis entre surcoût occasionné par le filtrage et réduction de l'espace de recherche.

# Méthode de travail

## Gestion de projet

Nous avons utilisé une méthode adaptative, la méthode « Scrum » vu en cours de génie logiciel. Celle-ci consiste à planifier le travail en constituant un « product backlog ». Les items seront évalués selon leurs valeurs, charges et risques.

Plus les items sont prioritaires, plus ils seront :

* Sous-divisés en sous-tâches pour une meilleure évaluation.
* Décrits sous diverses formes.

Les plannings effectués doivent être développés pendant une itération de durée fixe, nommée « sprint backlog ».

A la fin de chaque itération, un bilan est fait sur le travail effectué afin de le comparer à ce qui était planifié. Le bilan permet également de voir avec le client si le travail correspond à ses attentes. Le planning de la prochaine itération sera fait en fonction du résultat du bilan.

Malgré cette organisation, nous avons dû faire face à des contraintes et des obstacles qui ont ralenti l’avancement du projet. Nous avions mis beaucoup de temps à analyser et comprendre les objectifs. Il nous a fallu une grande période afin de se documenter sur le sujet, et de comprendre le fonctionnement des algorithmes.

Durant tout le projet, la difficulté principale était que l’on se mette d’accord sur la façon d’implémenter les algorithmes, les différences entre eux n’étaient pas claires pour nous au départ.

De plus, les périodes de révisons des examens du deuxième semestre ne nous ont pas permis de nous focaliser totalement sur le projet.

## Langage et outils de développement

### Langage de programmation

Nous avons utilisé le langage C pour la programmation ainsi que la génération des graphes. Ce langage était imposé dans le sujet du TER.

Il a fallu utiliser plusieurs bibliothèques propres au C :

Time.h et Math.h : Bibliothèques nécessaires pour pouvoir générer un nombre aléatoire.

Pour pouvoir tester les performances de chaque algorithme, on a dû utiliser plusieurs outils liés avec la plateforme éclipse :

* *Valgrind*:

Il s’agit d’un cadre d’instrumentation pour construire des outils d’analyse dynamique qui peuvent être utilisés pour profiler les applications en détail. Ces outils sont généralement utilisés pour détecter automatiquement de nombreux problèmes de gestion de mémoire et de filtrage. La suite *Valgrind* comprend également des fonctionnalités permettant de créer de nouveaux outils de profilage.

* *Oprofile* :

Cet outil est un profileur statistique pour les systèmes Linux. Ce logiciel est capable de profiler tout le code en cours d’exécution à faible surcharge.

* *Gprof* :

C’est un outil qui permet de savoir où un programme a dépensé son temps et quelles sont les fonctions qui ont fait un appel de méthodes durant leurs exécutions. Ces informations sont nécessaires afin de savoir quels sont les éléments les plus lents du programme.

### Travail collaboratif

Nous avons utilisé un gestionnaire de versions appelé Git, avec un dépôt central gratuit appelé GitHub. Notre choix s’est porté sur ce gestionnaire plutôt que *Hg*[[5]](#footnote-5) ou *SVN*[[6]](#footnote-6), car il offrait un meilleur compromis entre la puissance des fonctionnalités disponibles et la simplicité d’utilisation de l’outil. De plus, nous avions du en avoir recours lors du projet de Génie Logiciel.

*Lien du projet : https://github.com/AmadouK/TER.git*

Nous avons dû également communiquer via des mails entre nous et avec notre encadrant, afin de tester et de valider les tâches du sprint backlog.

# Travail réalisé

## Description du projet

Ce projet rentre dans le cadre de recherche national appelé TUPLES soutenu par l’Agence Nationale de la Recherche.

L’objectif est de :

* Générer un graphe aléatoire représentant les variables ainsi que les contraintes sur leurs valeurs du domaine.
* Implémenter et comparer l’algorithme du Foward Checking avec RFL. Les comparaisons seront sur le temps d’exécution des algorithmes, le pourcentage de CPU utilisé et le nombre d’affectations effectuées.
* Utiliser des heuristiques afin d’optimiser le choix des variables dans les algorithmes précédents.
* Etudier les comportements des algorithmes et comparer les résultats obtenus avec *FC* et *RFL* naïfs. En déduire si les heuristiques utilisées sont efficaces ou non.
* Proposer des améliorations pour optimiser le fonctionnement de chaque algorithme.

## Description des structures créés

Pour pouvoir représenter les graphes d’arc consistance, il a fallu créer deux structures :

* La première, nommée var\_domaine représente les valeurs qui doivent correspondre entre deux variables selon les contraintes. Elle est représentée par un tableau à deux dimensions de pointeurs d’entier. Si la case 0, 1 de l’instance de la structure est à 1, cela veut dire que la variable avec la valeur 0 peut être associée avec la variable de valeur 1.
* La seconde structure, appelée tab\_variable regroupe les liens entre les variables. Elle est représentée par un tableau à deux dimensions de pointeurs sur des var\_domaine. Ces pointeurs sont tous initialisés à Null à départ, et prennent l’adresse d’une instance de la structure var\_domaine s’il y a un arc entre deux variables.

*Structure tab\_variable*

0

1

2

…

n

…

n

0 1 … n

Null Null Null NULL

Null Null Null Null

Null Null Null Null

Null Null Null

Représente les variables 0 à n

Représente les variables 0 à n

Tableau à deux dimensions représentant les liens entre les variables

*Structure var\_domaine*

0

1

2

0 1 1

1 0 1

0 1 1

0 1 2

Représente le domaine de la première variable

Représente le domaine de la seconde variable

Tableau à deux dimensions représentant les correspondances entre les domaines des variables selon les contraintes

*Figure 3 : Relations entre les deux structures*

## Génération d’un graphe aléatoire

### Procédures pour générer le graphe

La génération d’un graphe aléatoire passe par deux fonctions :

* void genereGraphe(int nbSommet,int nbArrete,int nbDomaine,float sat)

Cette fonction aura pour but de créer une instance de la structure tab\_variable et de la configurer en fonction des paramètres nbSommet et nbArrete, qui sont le nombre de sommets et le nombre d’arrêtes dans le graphe. Les arrêtes seront générées aléatoirement, et les correspondances seront effectuées à l’aide de la fonction genereDomaine décrite ci-dessous.

* var\_domaine\* genereDomaine(float s,int nbDomaine,int var1,int var2)

Cette fonction a pour but de créer une instance de la structure var\_domaine et de la configurer en fonction des paramètres de la fonction. On aura donc un tableau à deux dimensions d’entiers de taille nbDomaine. Les correspondances des domaines seront générées aléatoirement, en ne dépassant pas le coefficient de satisfiabilité s.

### Modélisation du problème de 3-colorabilité avec la structure

On va représenter le graphe suivant dans la structure :

*Figure 4 : Exemple de graphe pour la 3-colorabilité*

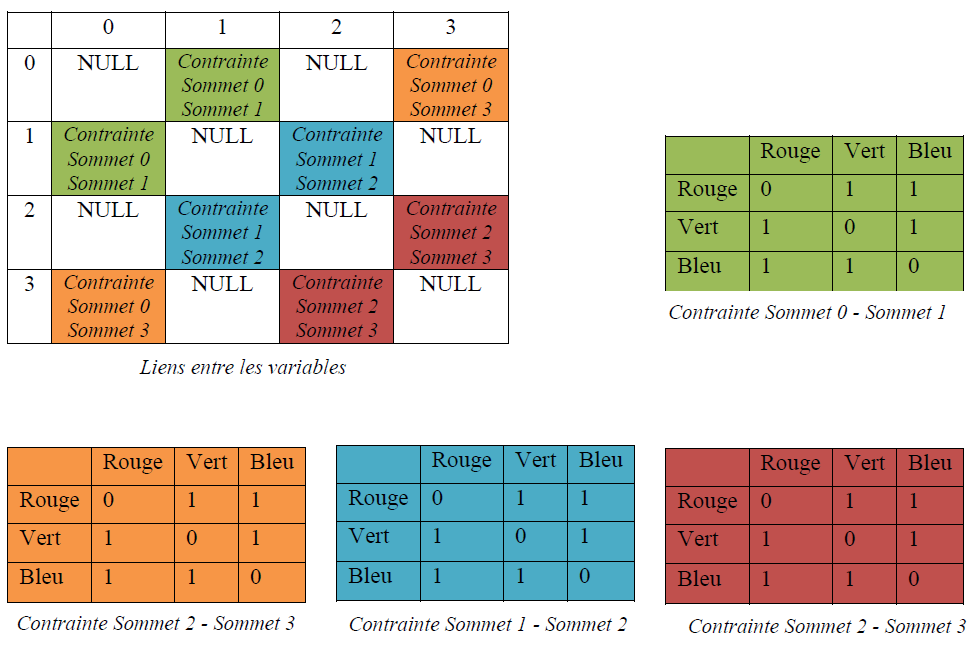
0

1

3

2

La figure ci-dessous représenter l’état de la structure après avoir implémenté le graphe avec les contraintes de la 3-colorabilité.



*Figure 5 : Génération du graphe obtenu avec les contraintes de la 3-colorabilité*

## Implémentation des algorithmes

### Backtrack

Il a fallu coder l’algorithme vu dans la partie 2.1.

Pseudo code de la fonction principale :

Fonction BT(variable : entier)

Début

Pour i de 0 à taille\_domaine faire

Si (verif\_Contrainte(variable,i))

Affectation de la variable à la valeur i

ok = BT(variable+1) ;

si ok

retourner 1

FinSi

FinSi

FinPour

On fait un backtrack

Fin

### Foward Checking

Il a fallu coder l’algorithme vu dans la partie 2.2.

Pseudo code des fonctions principales:

Fonction FC(variable : entier, values : entier\*\*)

Début

Pour i de 0 à taille\_domaine faire

x = values[variable][i]

Si (CF(values,variable,x))

Affectation de la variable à la valeur i

ok = FC(variable+1,values) ;

si ok

retourner 1

sinon

Récupération des valeurs du domaine avant affectation

FinSi

FinSi

FinPour

On fait un backtrack

Fin

Fonction CF(variable, x : entier,values : entier\*\*)

Début

Pour i de variable+1 à nb\_sommet faire

Si il y a un lien entre i et variable

Lien\_var = 1

Si variable peut avoir la valeur x en respectant les contraintes avec j

Ok = 1

FinSi

FinSi

Si ok = 0 et lien\_var = 0

Retourner 0

Sinon

Ok = 0

Lien\_var = 0

FinSi

FinPour

Retourner 1

Fin

### RFL

Il a fallu coder l’algorithme vu dans la partie 2.3.2.

Pseudo code de la fonction principale :

Fonction RFL(variable : integer)

Début

Récupération du domaine de toutes les variables

Pour i de 0 à taille\_domaine faire

Si variable peut avoir la valeur i

Affectation de la valeur i à la variable

AC3()

ok = RFL(variable+1)

Si ok

retourner 1

Sinon

Récupération des valeurs des valeurs de la structure avant affectation

finSi

finSi

finPour

backtrack

retourner 0

Fin

Pour que RFL puisse fonctionner, il a fallu implémenter un des algorithmes de filtrage. Nous en avons implémenté un.

Pseudo code des fonctions de « AC3 » :

Procédure AC3()

Début

L ← {(,), i ≠ j liées par une contrainte}

TantQue L ≠ ∅

Choisir et supprimer dans L un couple (,)

Si Revise(,) alors

L ← L ∪ {(,) / ∃ contrainte liant et}

finSi

finTantQue

Fin

Procédure revise(,)

Début

modification ← false

Pour chaque v ∈ faire

Si v’ ∈ | {→ v, → v’} est consistante alors

← \ {v}

modification ← vrai

finSi

finPour

retourner modification

Fin

## Intégration des heuristiques

Les algorithmes que nous venons d'étudier choisissent, à chaque étape, la prochaine variable à instancier parmi l'ensemble des variables qui ne sont pas encore instanciées. Ensuite, une fois la variable choisie, ils essayent de l'instancier avec les différentes valeurs de son domaine. Ces algorithmes ne disent rien sur l'ordre dans lequel on doit instancier les variables, ni sur l'ordre dans lequel on doit affecter les valeurs aux variables. Ces deux ordres peuvent changer considérablement l'efficacité de ces algorithmes.

Les heuristiques concernant l'ordre d'instanciation des valeurs sont généralement dépendantes de l'application considérée et difficilement généralisables. En revanche, il existe de nombreuses heuristiques d'ordre d'instanciation des variables qui permettent bien souvent d'accélérer considérablement la recherche. L'idée générale consiste à instancier en premier les variables les plus "critiques", c'est-à-dire celles qui interviennent dans beaucoup de contraintes et/ou qui ne peuvent prendre que très peu de valeurs.

L’ordre d’instanciation des valeurs peut être :

* Domaine le plus petit

On choisit les variables dont le domaine est le plus petit en premier dans l’algorithme. Si cette variable ne peut pas être instanciée, on se rendra compte plus rapidement que les contraintes ne sont pas satisfaisables.

* « First fail »

Les variables de l’algorithme sont choisies selon le calcul suivant : min

représente le domaine de la valeur i, et le nombre de liens avec les autres variables non affectées. On cherche avec cette heuristique à avoir un échec le plus rapidement possible.

## Graphes de comparaison

### Foward Checking

Le graphe ci-dessous représente le nombre d’affectations effectués sur le foward checking sans heuristiques sur le coefficient de satisfiabilité, nous avons décidé de partir avec un graphe de 100 sommets, 10 arcs, un domaine de 5 valeurs.

On peut remarquer que plus le coefficient se rapproche de 0,5, plus on aura de chance d’avoir plus d’affectations. On sera plus proche de 0 entre 0 et 0,5 et proche de 100 entre 0,5 et 1.

Le graphe suivant utilise est celui du foward checking avec l’heuristique du domaine le plus petit.

On remarque une différence entre les résultats du foward checking sans heuristique lorsque le nombre d’affectations, s’approche de 0,5. En effet, cette heuristique n’est efficace uniquement en cas d’échec et avec un coefficient de satisfiabilité élevé.

Le graphe suivant montre le nombre d’affectations effectué avec l’heuristique *« first fail ».*

Le nombre d’affectations est identique à ceux de l’heuristique du domaine le plus petit.

### RFL

Le graphe ci-dessous représente le nombre d’affectations effectués sur RFL sans heuristiques sur le coefficient de satisfiabilité, nous avons décidé de partir avec un graphe de 100 sommets, 10 arcs, un domaine de 5 valeurs.

On remarque que l’algorithme RFL fait un nombre d’affectation compris entre le nombre de variable du graphe. On constate la différence entre le foward checking et RFL.

Le graphe suivant utilise est celui de RFL avec l’heuristique du domaine le plus petit.

Les résultats montrent que cette heuristique ne marche pas sur le RFL. Le nombre d’affectations obtenus sont supérieurs à ceux obtenu dans le RFL sans heuristique.

Le graphe suivant montre le nombre d’affectations effectué avec l’heuristique *« first fail »*

Les résultats obtenus sont supérieur à ceux de l’heuristique du domaine le plus petit. On peut en conclure que pour l’algorithme RFL, il est inutile d’utiliser des heuristiques sur le choix des variables.

Conclusion

Ce projet nous a permis d’approfondir les connaissances que nous avions acquises lors des cours d’Intelligence Artificielle et de Complexité. Le travail en équipe a été bénéfique pour la réalisation du projet. Le temps que nous avions mis pour nous mettre d’accord sur des idées assimilées de manière différentes nous a donné un aperçu de l’organisation qu’il faudra pour réaliser des projets futurs au sein d’une équipe.

L’expérience de notre encadrant nous a beaucoup aidé pendant toute la durée du projet. Ses conseils, la documentation qu’il nous a fournis ainsi que les explications sur certains points lors des rendez-vous étaient précieux pour pouvoir coder les fonctions nécessaires.

Cette expérience a été un moyen de faire une découverte dans le monde de la recherche, et de pouvoir comparer avec les stages effectués durant notre cursus universitaire. Après réflexion, nous avons décidé de choisir un parcours professionnel après le Master 1 Informatique, vu l’effort et la patience requise dans le monde de la recherche.

Bibliographie

http://valgrind.org/

<http://oprofile.sourceforge.net/>

<https://eclipse.org/linuxtools/projectPages/gprof/>

<http://liris.cnrs.fr/csolnon/Site-PPC/session1/e-miage-ppc-sess1.htm>

<http://www.cril.univ-artois.fr/~lecoutre/research/publications/2004/JNPC2004Art.pdf>

1. *CSP : Problème à satisfaction de contraintes* [↑](#footnote-ref-1)
2. *Arc Consistency : (fr) Consistance d’arc* [↑](#footnote-ref-2)
3. *RFL : Real-full look-ahead* [↑](#footnote-ref-3)
4. *MAC : Maintaining Arc-Consistency* [↑](#footnote-ref-4)
5. *SVN : Apache Subversion, logiciel de gestion de versions* [↑](#footnote-ref-5)
6. *Hg : Mercurial, logiciel de gestion de version décentralisé* [↑](#footnote-ref-6)