



Rapport de stage

Stage réalisé par Régis JUBEAU à l'université de Mälardalen en Suède. (pour la 4^{ième} année de Microélectronique et Automatique de l'école d'ingénieur Polytech à l'université des sciences et des lettres de Montpellier)









Remerciements

Je tiens à remercier tout particulièrement Mme. Cristina SECELEANU, ma tutrice au sein de cette université pour son acceuil.

Je souhaite également remercier M. Tiberiu SECELEANU pour son accompagnement et ses renseignements.

Je voudrais aussi remercier M. Jakob DANILESSON pour son sujet de projet, son soutien, son aide et ses conseils.







Table des matières

Remerciements	2
Introduction	
1 – Déroulement de mon projet	5
1.1 - Introduction du projet de stage	5
1.2 - Réalisation de l'interface graphique	7
1.2.1 - Lecture des fichiers .txt sélectionnés par l'utilisateur	7
1.2.2 - Analyse des caches misses en fonction des fichiers	8
1.2.3 - Affichage des partitions et des graphiques en fonction de l'application	11
1.2.4 - Le reste de l'interface graphique	12
1.3 - Conclusion pour l'interface graphique	13
2 – Assemblage des projets	14
3 – Conclusion	15







Introduction

Mälardalen Högskola est une université suédoise dont un campus est situé à Västerås. Elle compte six spécialisations de recherche différentes : sciences de l'éducation et mathématiques, systèmes embarqués, énergie du futur, santé et bien-être, économie et gestion industries, et innovation et réalisation de produits. L'université de Mälardalen est organisée en quatre écoles dont l'école d'innovation, de design et d'ingénierie où j'effectue mon stage.

Mon stage avait pour principal objectif de créer une interface graphique sur le framework Qt codé en C++ qui serait additionnée à une autre interface graphique (réalisée par M. LAURENDEAU) pour compléter le projet global. L'interface graphique globale doit rendre meilleure l'éxécution, l'affichage et la comparaison des caches misses par rapport au différents évènements de PAPI (Performance Application Programming Interface) que vous découvrirez au fur et à mesure de ce rapport. Vous suivrez également le chemin de ces trois mois passés auprès de chercheurs et thésards. Vous découvrirez tous les problèmes et résolutions prises pour finir cette interface graphique et l'assemblage.







1 – Déroulement de mon projet

1.1 - Introduction du projet de stage

Pour introduire mon sujet de stage, voici une mise en situation. Les systèmes multicœurs sont en train de devenir la norme de fait dans les domaines de l'informatique commercial et de l'informatique embarquée. Les processeurs multicœurs présentent une plus grande capacité de calcul tout en offrant une diminution de la taille et du poids par rapport à leurs prédécesseurs à simple cœur. Les processeurs multicœurs permettent la parallélisation au niveau des applications et du système et l'exécution simultanée de différentes applications sur différents cœurs. Les systèmes multicœurs mettent souvent en œuvre une structure interne de ressources de déchiquetage pour augmenter la vitesse de communication entre les cœurs. Parmi les exemples de ressources partagées, citons les Translation lookaside Buffers (TLB), le Last Level Cache (LLC), le bus mémoire et le périphérique d'E/S à usage général. Voici un de structure de ressources partagées typique dans la figure ci-dessous :

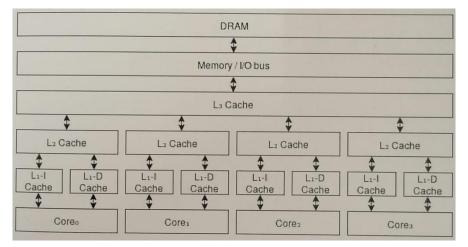


Figure 1: Exemple d'un système à 4 cores

Mon sujet de stage a été proposé par M. Danielsson, un doctorant de l'université de Mälardalen depuis Août 2016. Sa thèse, « <u>Automatic Characterization and Mitigation of Shared-Resource Contention in Multi-core Systems</u> », se concentrait sur les systèmes multi-cœurs homogènes qui utilisent deux ou plusieurs cœurs pour exécuter des applications. Elle ciblait le Last-level cache (LLC), un terme englobant tous les caches situés en dernier dans la hiérarchie de la mémoire. Le LLC est le plus souvent physiquement partagé entre les différents cœurs d'un processeur et relie la connexion entre le cache local d'un processeur et le bus mémoire. Il est particulièrement sujet à la contestation de ressources partagées puisqu'il est partagé entre différents cœurs. Lors de sa thèse, sa recherche visait principalement la contestation des ressources partagées, qui se produit en raison de l'utilisation simultanée de plusieurs cœurs.

Lors de sa thèse, il a réalisé un code où il a utilisé PAPI (Performance Application Programming Interface) qui est une interface de programmation portable permettant d'accéder aux compteurs matériels spécifiques aux microprocesseurs modernes. PAPI est utilisé pour collecter des informations de bas niveau, telles que le nombre d'opérations en virgule flottante par seconde, le nombre de cache







misses durant l'exécution d'un code, etc. Cette méthodologie possède une liste d'évènements et en voici un exemple que j'ai pu rencontrer lors de mon stage :

```
PAPI Preset Events
                                                                                                             Level
Level
Level
                                                                                                                                 1 data cache misses
1 instruction cache
                                                                                                                                 2 data cache misses
2 instruction cache misses
3 data cache misses
                                                                                                              Level
Level
                                                                        No
No
Yes
                                                                                                                                        instruction cache misses
                                                                         No
No
No
                                                                                                              Requests for exclusive access to shared cache line
Requests for exclusive access to clean cache line
Requests for cache line invalidation
Requests for cache line intervention
                                                                                             No
No
No
No
No
No
No
No
Yes
                                                                                                                                  3 load misses
                                                                                                            Level 3 store misses
Cycles branch units are idle
Cycles integer units are idle
Cycles floating point units are idle
Cycles load/store units are idle
Data translation lookaside buffer misses
Instruction translation lookaside buffer misses
Total translation lookaside buffer misses
                                                                         No
No
No
No
Yes
                                                                        Yes
No
Yes
Yes
                                                                                                                                 1 store misses
2 load misses
2 store misses
                                                                                                             Branch target address cache misses
Data prefetch cache misses
Level 3 data cache hits
Translation lookaside buffer shootdowns
Failed store conditional instructions
                                                                        No
No
```

Figure 2: Exemple d'évènement de PAPI

Lors de l'éxécution de ce code, en fonction des counters choisis, un fichier .txt était retourné, composé de 3 colonnes : un index, le nombre d'opérations par seconde et le nombre de cache misses sous cette forme :

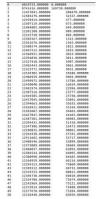


Figure 3: Exemple de fichier .txt

Donc, l'objectif de mon stage a été de réaliser une interface graphique sur le framework Qt codée en C++. Cette interface graphique devait lire des fichiers .txt, et les analyser. Le but étant d'avoir des fichiers du même évènement mais réalisé avec des applications différentes. Et affecter à chacune des applications, une nombre de partition en fonction du nombre de caches misses.







1.2 - Réalisation de l'interface graphique

Une image de l'interface graphique sera mise à la fin de cette partie (figure 18).

Pour les exemples, ce sera toujours le même évènement « PAPI_L2_DCA » avec les mêmes applications (« FAST », « SIFT », « SUSAN ») pour voir l'évolution au fur et à mesure de l'explication.

1.2.1 - Lecture des fichiers .txt sélectionnés par l'utilisateur

Pour le projet, il fallait que ce soit l'utilisateur qui ait le choix des fichiers .txt qu'il voulait analyser. Pour cela, j'ai crée un bouton qui nous amène dans les fichiers de l'ordinateur. Ensuite, l'utilisateur n'a plus qu'à sélectionner ce qu'il souhaite.

Figure 4: Fonction du bouton "OPEN"

On peut voir sur cette image, le code pour le bouton. Grâce à Qt et ses fonctions, dès que l'utilisateur cliquera sur le bouton, il rentrera dans cette fonction. Cette fonction récupère le nom du fichier sélectionné qu'elle enregistrera en l'ajoutant dans le vecteur « file_path ». Cette fonction vérifiera aussi le nombre de fichier ouvert par rapport aux nombres de core que possède l'ordinateur d'où l'indicateur de fichier ouvert. En cas d'un nombre supérieur, une alerte s'ouvrira.

Pour la partie lecture du fichier, c'est la fonction « create_vector » (Annexe 1) qui s'en occupe. À savoir, les fichiers .txt sont écrits sous forme de colonnes et elles sont toutes séparées par le caractère « tabulation ». La fonction récupère le vecteur contenant le nom de chacun des fichiers sélectionné et les ouvre un par un. Une fois ouvert, la fonction lit chaque valeur (double) et l'ajoute dans un vecteur qui lui correspond. Puis ses vecteurs sont eux-même ajoutés dans un autre vecteur ce qui me permet d'avoir dans un seul vecteur toutes les valeurs du nombre d'instructions ou encore le nombre de caches misses. Voici le résultat de cette fonction en image :

```
Nombre de caches misses:

Qvector(Qvector(0, 9987, 9957, 25591, 184855, 15453, 114636, 184970, 298776, 309192, 19874, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 11872, 118
```

Figure 5: Résultat de la fonction "create_vector"

En n'ouvrant que trois fichiers, on voit qu'un vecteur en comprend trois.







1.2.2 - Analyse des caches misses en fonction des fichiers

Pour cette partie, je devais partager un nombre de partitions (16 partitions) en fonction du nombre de caches misses de chaque application faite sur un évènement. Pour cela, j'ai procédé par étape.

Tout d'abord, je devais réaliser une fonction qui trouvait les segmentations de mes vecteurs. Voici un schéma approximatif et représentatif de ma fonction :



Figure 6: Schéma de la segmentation des fichiers

Pour la créer, j'ai créé une autre fonction qui me renvoyait la moyenne globale du vecteur que je mettais en entrée, et j'analysais par la suite, chaque valeur du segment par rapport à celle-ci. Si une des valeurs était inférieur à 80 % de la moyenne, alors la fonction crée une segmentation. Cette fonction est nommée « find_segmentation » (Annexe 2).

Figure 7: Fonction "find_segmentation"

Qvector(Qvector(0, 1, 2, 4, 23, 24, 25, 26, 27), Qvector(0, 1, 2, 37, 38, 39, 41, 42, 43, 44, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77), Qvector(0, 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 22))

Figure 8: Résultat de la fonction "find_segmentation"

D'après les résultats, on peut voir qu'il y a des segmentations consécutives trop proches que nous allons supprimer par la suite.







Pour chaque vecteur de segmentations, je vérifiais que les segmentations n'étaient pas trop proches ni trop éloignées pour éviter d'avoir un nombre trop important ou peu important de segmentations. Pour cela, je vérifiais si la différence de chaque valeur consécutive était supérieur à 10 ou supérieure à 20. Dans ces cas là, soit je supprimais la valeur inférieure pour ne former plus qu'une seule segmentation, soit j'ajoutais une segmentation à environ +15 par rapport à la plus petite valeur.

```
QVector(QVector(4, 19, 27), QVector(2, 17, 32, 47, 62, 77), QVector(14, 22))
```

Figure 9: Résultat de la fonction "correction_segm"

D'après le résultat, nous sommes passés pour le premier vecteur de 9 segmentations à 3, pour le deuxième de 15 segmentations à 2 et pour le troisième de 40 à 6. Dans ce résultat, on peut voir pour le premier vecteur, qu'il y a la segmentation « 19 » qui n'apparaissait pas de base. Elle a été ajoutée car la différence entre 27 et 4 était supérieure à 20. Ce qui permet aux vecteurs de segmentations d'être plus homogène.

Ensuite, ces fichiers .txt doivent être comparés entre eux, pour cela, j'ai décidé de ne créer qu'un seul vecteur de segmentations réunissant tous les autres comme le montre l'exemple suivant :

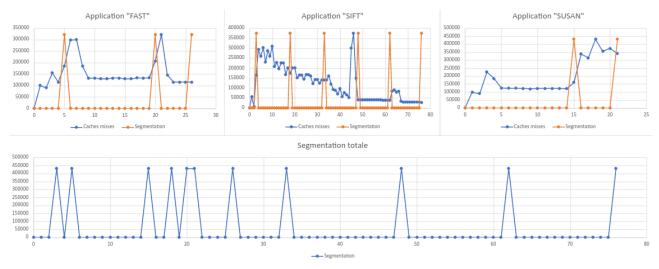


Figure 10: Schéma du rassemblement des différentes segmentation en fonction des fichiers

Sur le schéma du dessous, toutes les segmentations ont été ajoutées pour ne former qu'un vecteur. Bien sûr, pour l'application « FAST » et l'application « SUSAN », l'index ne va pas à plus de 30, mais ici ce sont des graphiques d'application réalisés auparavant par Mr. Danielsson qui n'ont pas la même durée.

Pour créer ce vecteur, j'ai réalisé une fonction, qui ajoute les segmentations de tous les autres vecteurs à un seul vecteur, nommée « create_segmentation_and_avg_total » (Annexe 14). De plus, elle le trie et supprime les doublons comme le résultat suivant :

```
Vecteurs de segmentation par fichier:

QVector(QVector(4, 19, 27), QVector(2, 17, 32, 47, 62, 77), QVector(14, 22))

Vecteur de toutes les segmentations:

QVector(2, 4, 14, 17, 19, 22, 27, 32, 47, 62, 77)
```

Figure 11: Résultat dde la première partie de la fonction "create_segmentation_and_avg_total"







On voit donc que le dernier vecteur est la somme des trois autres vecteurs.

Cette fonction a aussi une autre utilité. Elle calcule la moyenne du nombre de caches misses entre chaque segmentation de ce nouveau vecteur pour pouvoir les analyser après. Voici le résultat :

```
Vecteur de toutes les segmentations:
QVector(2, 4, 14, 17, 19, 22, 27, 32, 47, 62, 77)

Noyenne entre chaque segmentations:
QVector(9Vector(99719, 245016, 1.73774e+06, 390984, 265019, 662435, 604443, 0, 0, 0, 0), QVector(55820, 173255, 2.57518e+06, 620561, 375834, 552447, 807723, 691127, 1.97961e+06, 715909, 675761), QVector(98897, 315138, 1.28823e+06, 619907, 746477, 1.07012e+06, 0, 0, 0, 0, 0))
```

Figure 12: Résultqt de la deuxième artie de la fonction "create_segmentation_and_avg_total"

Dans ce résultat, on peut voir que certain vecteur sde moyenne possèdent des « 0 ». Ceci est dû à la différence de taille des fichiers et donc le programme adapte les vecteurs. En effet, si une application finit plus tôt que les autres, il est logique que son nombre de caches misses soit à 0 et donc sa moyenne aussi.

Enfin, toutes ces étapes intermédiaires réalisées, j'ai pu commencer à créer une fonction qui calculait le nombre de partition en fonction du nombre de caches misses. Pour illustrer cela, voici un schéma :

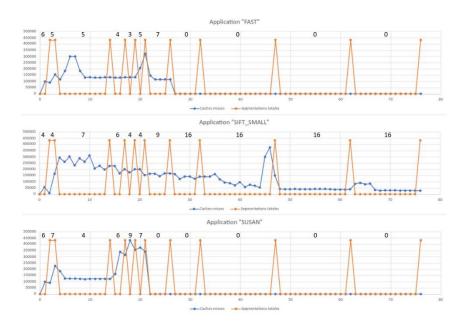


Figure 13: Schéma du partitionnage des applications en fonction de leur nombre de caches misses respectif

Les nombres situés en haut de chaque courbe représentent le nombre de partitions que l'on pourrait accorder à ces différentes applications selon le nombre de caches misses. Mais pour réaliser ceci, j'ai créé une fonction nommée « create_partition » (Annexe 15). Cette fonction fait une somme de toutes les moyennes entre deux segmentations, puis réalise une règle de trois et multiplie par le nombre de partitions disponibles (ici 16 partitions). Ceci réalisé, les valeurs de partitions sont mises dans leur vecteur respectif, et ses vecteurs sont mis dans le vecteur des partitions comme le montre le résultat suivant :

```
Vecteur des partitions en fonction du fichier :

QVector(QVector(6, 5, 5, 4, 3, 5, 7, 0, 0, 0, 0), QVector(4, 4, 7, 6, 4, 4, 9, 16, 16, 16, 16), QVector(6, 7, 4, 6, 9, 7, 0, 0, 0, 0))
```

Figure 14: Résultat de la fonction "create_partition"







On remarque donc que le partage correspond à peu près avec les moyennes obtenues avant. De plus, les valeurs sont entières, ceci est dû à une fonction que j'ai créée car une partition est forcément entière.

Pour être sûr que mes résultats soient corrects, j'ai fait les mêmes calculs avec les mêmes fichiers sur l'application « excel » et en voici les résultats :

Segmentation	AVG "FAST"	AVG "SIFT_SMALL"	AVG "SUSAN"	SUM TOTAL	PART "FAST"	PART "SIFT_SMALL"	PART "SUSAN"
(0-2)	190172	63782	188444	442398	7	2	7
(2-4)	269199	458211	410486	1137896	4	6	6
(4-14)	1755003	2509094	1408930	5673027	5	7	4
(14 - 17)	391933	595156	813869	1800958	3	5	7
(17 - 19)	265350	374567	787253	1427170	3	4	9
(19 - 22)	529254	352286	714683	1596223	5	4	7
(22 - 27)	604443	807723	0	1412166	7	9	0
(27 - 32)	0	831829	0	831829	0	16	0
(32 - 47)	0	1988583	0	1988583	0	16	0
(47 - 62)	0	604802	0	604802	0	16	0
(62 - 77)	0	637195	0	637195	0	16	0

Figure 15: Résultat du partitionnage des applications en fonction de leur nombre de caches misses respectif sur "Excel"

On remarque que les résultats sont pour la plupart les mêmes entre « excel » et mon programme. J'en ai donc déduit que mes calculs étaient bons.

1.2.3 - Affichage des partitions et des graphiques en fonction de l'application

Pour la partie affichage qui est la partie la plus importante de mon interface graphique, j'ai procédé de la même manière que ce soit pour l'affichage des partitions que pour l'affichage des graphiques.

Tout d'abord, pour afficher les résultats sur l'interface graphique, j'avais pensé à faire comme sur la figure 13, c'est-à-dire afficher les valeur au-dessus de leur courbe respective. Ni parvenant pas, j'ai eu l'idée de créer un tableau où j'afficherai toutes les partitions en fonction de leur application. Pour cela, j'ai créé un tableau où la première ligne correspondait aux segmentations, et les autres lignes se créaient en fonction du nombre de fichiers ouverts. Ensuite, je n'avais plus qu'à ajouter les valeurs à l'intérieur de celui-ci, comme le montre le résultat en dessous :



Figure 16: Affichage du partitionnage sur l'interface graphique







Ensuite, pour afficher les graphiques sur l'interface graphique, j'ai également créé un tableau qui ajoute autant de graphique que de fichiers ouverts. Étant donné qu'ils sont dans un tableau, il suffit juste de scroller. Il me suffisait juste de régler les paramètres de Qt pour la partie « plot » et de lui envoyer les valeurs voulues pour obtenir ce résultat :

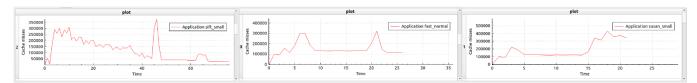


Figure 17: Affichage des graphique des appplications "SIFT_SMALL", "FAST" et "SUSAN" sur l'interface graphique

Nous pouvons zoomer et déplacer le graphique en fonction de ce que l'on veut voir comme j'ai pu le faire pour certains de ces graphiques.

1.2.4 - Le reste de l'interface graphique

En plus de toutes ces fonctionnalités, j'ai deux boutons qui me servent, un à lancer le programme une fois que l'utilisateur a sélectionné ses fichiers qu'il veut les analyser, et l'autre, pour tout nettoyer et sélectionner de nouveau fichier pour éviter de devoir fermer l'application et la relancer à chaque fois.

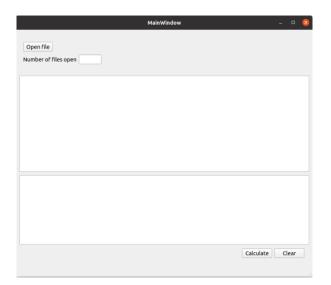






1.3 - Conclusion pour l'interface graphique

Pour conclure, l'interface graphique est finie et j'ai atteint les objectifs mis en place avec ma tutrice Mme. Seceleanu et le doctorant M. Danielsson. Plusieurs problèmes sont survenus lors du projet mais ont été assez vite régler grâce au débuggage de celui-ci. Pour voir le résultat final, je vous mets endessous des images de l'interface graphique :



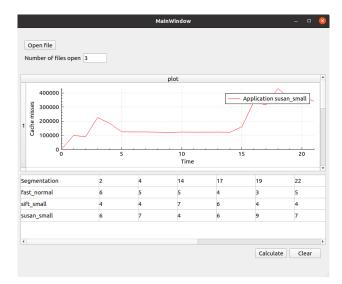


Figure 18: Interface graphique "non utilisée" et "utilisée"







2 – Assemblage des projets

Après la réalisation de nos deux projets respectifs, nous avons dû les rassembler pour ne former qu'une seule interface graphique. Pour cet assemblage, j'ai dû créer une classe avec toutes mes fonctions et l'interface graphique par rapport à celle-ci. Une fois la classe réalisé, il fallait que je relie le bouton « Analysis » de l'interface graphique de M. LAURENDEAU à cette classse. Pour cela, j'ai créé une fonction qui lorsque l'on clique sur ce bouton, mon interface graphique s'affiche sur une nouvelle fenêtre. Puis après, l'utilisateur n'a plus qu'à utiliser l'interface comme indiqué auparavant. Lorsque l'application est lancée, nous pouvons accéder à mon interface graphique en permanence sauf quand la fonction « Run » de M. LAURENDEAU est en marche. Ceci est un choix entre lui et moi pour une question de pratique par rapport à l'interface graphique. La fonction « Run » permet d'afficher en temps réel, les valeurs renvoyées par les évènements PAPI et donc par sécurité, aucune autre fonction ne peut être réalisée lors de l'éxécution de celle-ci jusqu'à temps qu'elle se termine, ou si vous appuyez sur le bouton « Stop ».

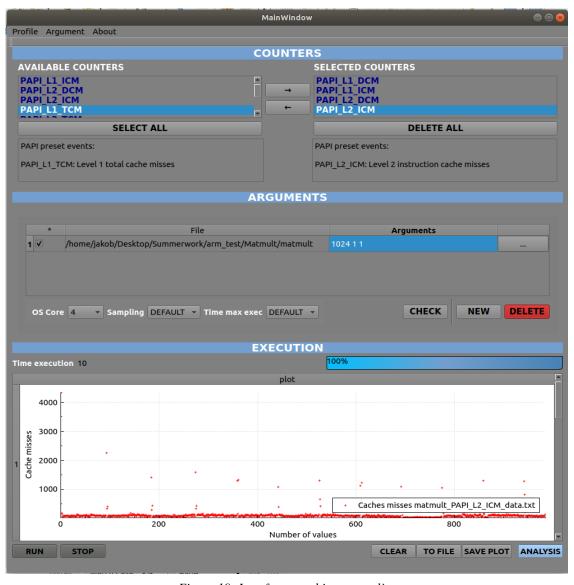


Figure 19: Interface graphique complète







Et lorsque l'on clique sur « Analysis », nous avons cette fenêtre qui s'ouvre :

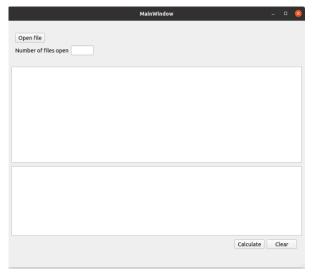


Figure 20: Partie "Analysis" de l'interface graphique

3 – Conclusion

Ce stage, avec le fait de voyager, d'apprendre et de se perfectionner en anglais, fut une excellente expérience. Le début fut compliqué entre le sujet, la compréhension et la mise en route du programme. Mais au final, j'ai su contrer chaque difficulté pour aujourd'hui rendre un bon résultat qui a été accepté par ma tutrice Mme. Seceleanu et le doctorant M. Danielsson. Ce projet va permettre à notre doctorant et nos tuteurs de visualiser et de comparer en direct les caches misses de l'ordinateur qui seront utilisés au cours des applications.