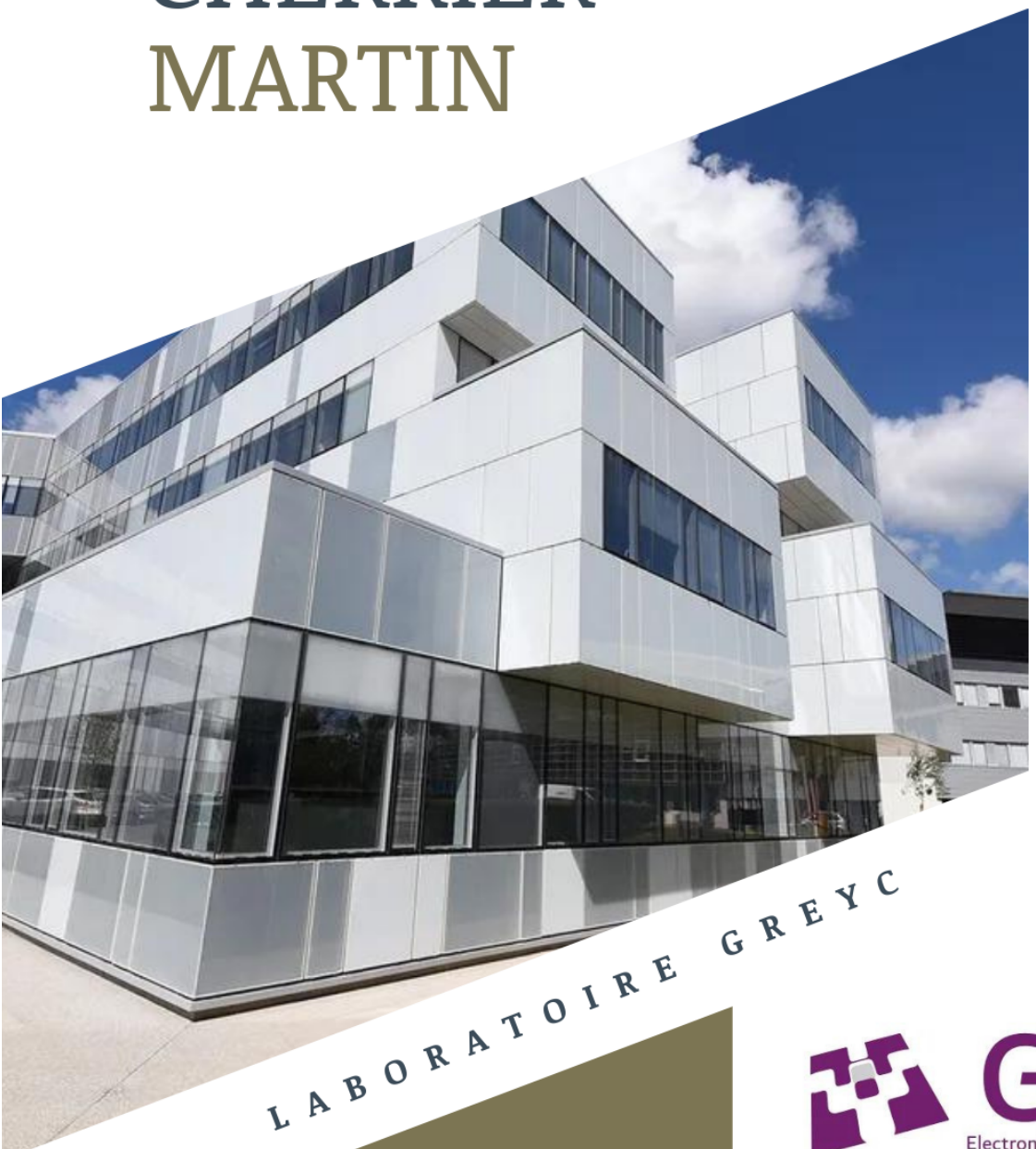
RAPPORT DE STAGE

IUT GRAND OUEST NORMANDIE

UNIVERSITÉ CAEN NORMANDIE

20 FÉVRIER 2023 - 28 AVRIL 2023

CHERRIER MARTIN



Sous la tutelle de :

- Eddy Godelle,
  Professeur Référent à l'IUT
- Christophe Rosenberger
  Maître de stage

LABORATOIRE GREYC

GREYC
Electronics and Computer Science Laboratory

6 Boulevard du Maréchal Juin
Bâtiment F
CS 45053, 14050 CAEN cedex 4

Remerciements

Tout d'abord, je tiens à remercier M. Christophe Rosenberger, mon maître de stage, pour m'avoir accordé sa confiance sur cet important projet et pour m'avoir confié différentes missions qui furent intéressantes, qui m'apportèrent de précieuses connaissances.

Je tiens ensuite à remercier l'équipe SAFE m'ayant accueillie chaleureusement en son sein. Et plus particulièrement Hugo Jean, Adrien Dubettier ainsi que Simon Cardoso, principaux acteurs du projet G'DIP qui ont pris le temps de me présenter chacun leur tour la partie du projet les concernant. Ce qui m'a permis d'améliorer ma compréhension globale de ce dernier.

Je tiens aussi à remercier M. Ostanciaux et Mme. Fior pour leur aide dans les différentes tâches administratives.

Je remercie également M. Godelle, mon tuteur et référant à l'IUT, de m'avoir suivi durant les 10 semaines de stage ainsi que d'être venu me rendre visite durant le stage.

Enfin, je remercie l'ensemble de l'équipe enseignante du BUT Réseaux et Télécommunications pour m'avoir enseigné toutes les connaissances nécessaires à la réalisation du stage.

Sommaire

Introduction

Lors de la deuxième année de BUT réseaux et télécommunications, il nous est demandé de réaliser un stage d'insertion en entreprise d'une durée comprise entre 8 et 10 semaines. J'ai réalisé ce stage dans le **G**roupe de **R**echerche en **I**nformatique, **I**mage, Automatique et **I**nstrumentation de **Caen,** aussi appelé GREYC et plus précisément dans l'équipe SAFE, **S**écurité, **A**rchitecture, **F**orensique et biom**É**trie. J'ai choisi le GREYC dû au fait que c'est un groupe de recherche en informatique influent sur Caen et c'est aussi dû au fait que je trouvais intéressant de découvrir le monde laborantin.

Durant ce stage, j'ai pu participer à la vie quotidienne de l'équipe en aidant sur différents projets et plus particulièrement sur le projet G'DIP, un projet ayant pour but d'aider les enquêtes criminelles, mais aussi d'aider l'analyse d'archives numériques. Ce projet est réalisé en partenariat avec la gendarmerie de Caen ainsi que l'IMEC, l'Institut Mémoire de l'édition contemporaine. Ainsi que d'autres projets et tâches annexes liés à l'équipe.

Mon rôle dans le projet était de faire la partie matérielle, c'est-à-dire l'extraction de données provenant d'ordinateur ou de téléphone. Pour cela, il me faudra comprendre comment la machine FRED fonctionne, comment pourrons-nous nous en servir afin d'avancer dans le projet et d'en faire des documentations pour faciliter la compréhension de tous. Une fois cette partie finie, j'ai pu faire des images de disques ainsi que des analyses. De plus, j'ai aussi pu effectuer d'autres tâches.

Ce rapport résume mon parcours et mes différentes activités durant ces 10 semaines. Ce rapport sera centré sur des notions de forensique ainsi que des notions de programmation et sera basé sur différentes parties, à savoir, en premier lieu, la présentation du laboratoire et de ses tutelles ainsi que du projet sur lequel j'ai travaillé. La partie suivante sera sur mes activités quotidiennes au sein de l'équipe et du projet. Puis, je finirai par faire un bilan du stage en revenant sur ce que j'ai pu faire et ce que j'ai appris.

1 - Présentation du laboratoire et du projet

1.1 Les différentes tutelles

L'**É**cole publique **N**ationale **S**upérieure d'**I**ngénieurs de Caen est une école d'ingénieurs récente, elle fait suite au changement de nom de l'ISMRA, l'**I**nstitut des **S**ciences de la **M**atière et du **R**ayonnement. Ce même institut venant d'anciennes écoles et autres structures d'enseignement plus anciennes. La plus ancienne datant de 1911 ou 1912 s'appelant l'**I**nstitut des **S**ciences **A**ppliquées de Caen. L'ENSI est le site accueillant le laboratoire GREYC et c'est aussi l'une de ses trois tutelles.

Le **C**entre **N**ational de la **R**echerche **S**cientifique est le principal ainsi que le plus grand organisme de la recherche scientifique. Fondé en 1939, ce dernier aide les différents laboratoires comme le GREYC. Le CNRS est l'une des trois tutelles.

La dernière des trois tutelles est l'Université de Caen, fondée en 1432 suite à un projet du Pape Martin V, c'est la plus grande université de Normandie à l'heure actuelle. Cette université est composée de différentes formations ainsi que de plusieurs laboratoires sous tutelles. C'est aussi l'université dans laquelle j'évolue en tant qu'étudiant du Campus 3 et du BUT Réseaux et Télécommunications.

1.2 Le laboratoire GREYC

Le **G**roupe de **R**echerche en **I**nformatique, **I**mage, automatique et **I**nstrumentation de **C**aen est un groupe de recherche formé suite au regroupement de plusieurs enseignants-chercheurs exerçant dans le domaine de l'informatique ainsi que celui de l'électronique. Dans les années suivantes, plusieurs autres groupes ont rejoint le premier cercle et ont ainsi formé le GREYC. Depuis le début des années 2000, le groupe est une unité de recherche liée à l'ENSI, au CNRS ainsi qu'à l'Université de Caen.

Ce groupe de recherche est maintenant composé de plusieurs équipes :

- L'équipe SAFE (**S**écurité, **A**rchitecture, **F**orensique, Biom**É**trie), faisant des recherches dans trois sujets complémentaires, la biométrie, l'architecture et les modèles de sécurité ainsi que les sciences de l'investigation (Forensique). C'est dans cette équipe que j'ai évolué durant mon stage.

- L'équipe AMACC (**A**lgorithmes, **M**odèles de calcul, **A**léa, **C**ombinatoire, **C**omplexité), faisant des recherches en informatique mathématique et plus précisément à deux concepts de cette matière, la complexité et à l'algorithmique.

- L'équipe CODAG (**C**ontraintes, **O**ntologies, **D**onnées, **A**nnotations, **G**raphes), faisant des recherches sur la chaîne de traitements des données sous différentes contraintes, aussi bien légales que physiques.

- L'équipe ÉLECTRONIQUE, faisant des recherches sur les composants électroniques ainsi que des capteurs à haute sensibilité afin d'en améliorer les performances.

- L'équipe IMAGE, faisant des recherches sur le développement de nouvelles méthodes de traitement et d'analyse de signaux ou d'images en partenariat avec des centres de recherche en imagerie biomédicale.

- L'équipe MAD (**M**odèles, **A**gents, **D**écision), faisant des recherches sur les intelligences artificielles et plus particulièrement sur le raisonnement et la représentation de connaissances ainsi que la planification sous incertitude

Chacune des équipes gère différents projets et est composée d'enseignants-chercheurs ou d'étudiants réalisant des thèses, appelés non-permanents. Durant mon stage, j'ai pu participer à un projet créé par l'équipe SAFE. Ce projet est nommé G'DIP pour **G**REYC **D**igital **I**nvestigation **P**latform.

.

1.3 Le Projet G'DIP

Durant les premières semaines de mon stage, j'ai eu l'occasion de participer au projet introduit précédemment, le projet G'DIP. Le projet **G**REYC **D**igital **I**nvestigation **P**latform est l'un des principaux projets actuels de l'équipe SAFE, en partenariat avec l'**I**nstitut **M**émoires de l'**É**dition **C**ontemporaine ainsi que la gendarmerie de Caen. Le but de ce projet est de développer une plateforme permettant l'analyse de disques durs dans un contexte d'enquêtes de police ainsi que d'archives de documents.

Ce projet est divisé en plusieurs parties, une première partie utilisant des intelligences artificielles en utilisant différentes bases de données pour les entrainer, une seconde s'occupant des scripts qui serviront de filtres liés à l'interface du site web, la troisième partie étant l'interface web servant à visionner les résultats des analyses d'images en utilisant différents graphiques. La dernière et quatrième partie étant celle dont je me suis occupé qui consiste en différents tests pour créer des images de disques durs en utilisant la machine FRED, que je présenterai ci-dessous, ainsi que d'analyser en utilisant une plateforme spécialisée, *Autopsy*, ces images générées.

De plus, j'ai eu à réaliser un inventaire de la machine, des différents câbles et des adaptateurs fournis avec la machine ainsi que de rédiger des méthodologies sur les différentes tâches que j'ai réalisées et rédiger des documentations sur la machine afin que d'autres personnes puissent s'en servir et répéter les actions que j'ai faîtes.

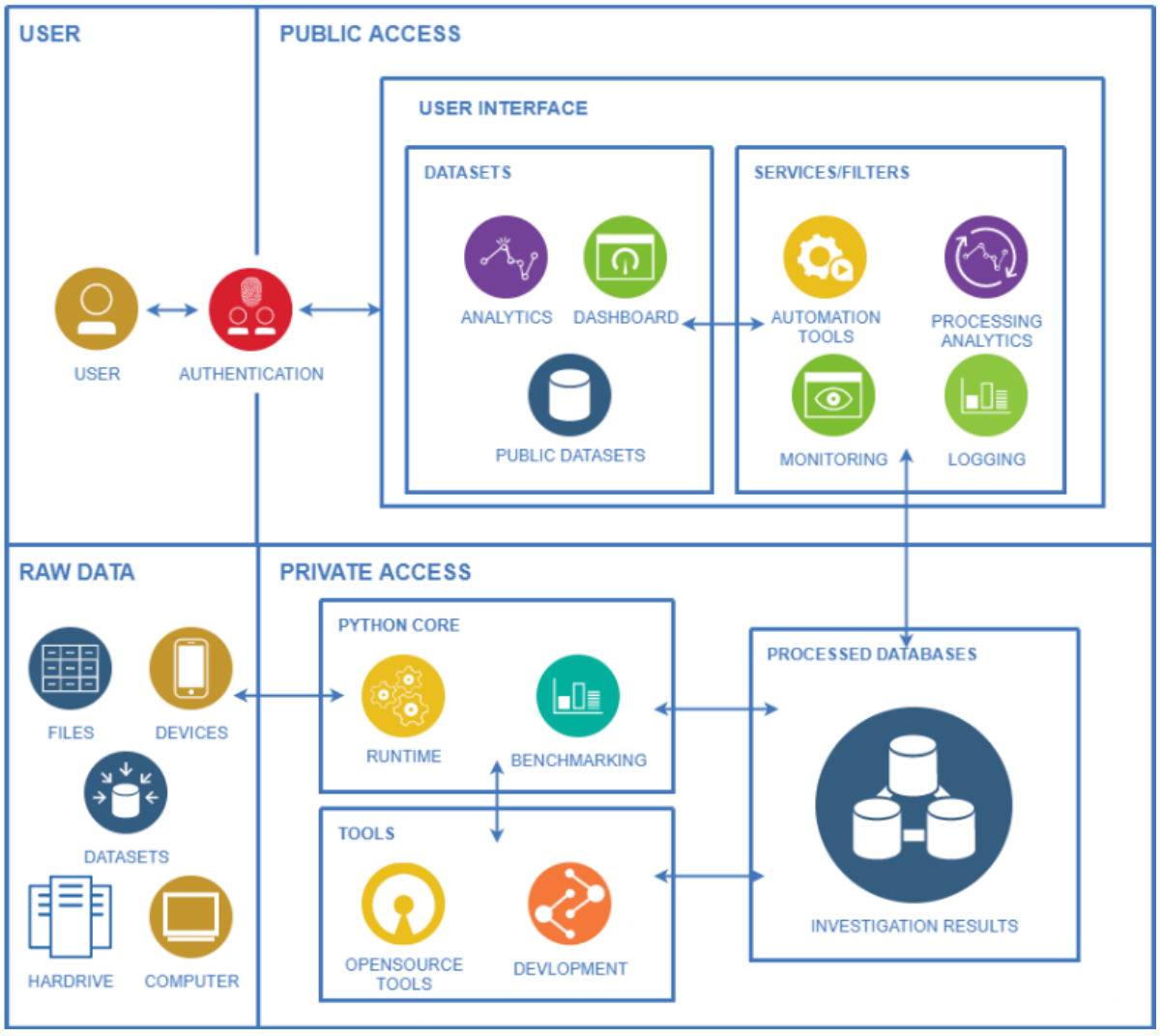


Diagramme officiel du projet G'DIP

2 - Développement

2.1 Le Projet G'DIP

2.1.1 La machine FRED

Créée par Digital Intelligence, FRED pour **Forensique** **R**ecovery **E**vidence **D**evice est un ordinateur servant dans le domaine de la forensique, composé d'une baie UltraBay 4D, d'une baie permettant de ranger des cartes SD ou des cartes flash ainsi que de différents racks permettant le stockage de données. Certains racks de stockage de données sont Hot Swap, c'est-à-dire qu'ils peuvent être enlevés et modifiés même si la baie est allumée.

Le reste des composants est similaire à un ordinateur normal.

Les racks ne sont que des supports accueillant des disques durs ou des SSD.

Il existe d'autres machines ayant la même utilité que celle-ci, mais développées par d'autres entreprises. Telles que la station venant de l'entreprise Cellebrite.

Certaines parties telles que l'Ultra Bay 4D que je présente ci-dessous, peut être remplacée par tout type de bloqueur d'écriture, il en existe différents types. Certains sont portables et d'autres plus compacts.

L'Ultra Bay 4D est l'élément principal de cette machine, elle est composée de différents ports permettant le branchement de différents équipements stockant des données ainsi que d'un petit écran tactile affichant les informations sur les équipements reliés. Cette même baie permet de bloquer l'écriture sur ces disques, rendant l'utilisation de ces derniers plus sécurisée en cas d'enquêtes durant lesquelles les données ne doivent pas être altérées.

Baie 4D présente sur la machine.

Cette machine était fournie avec différents logiciels permettant l'analyse et la création d'image via la baie citée précédemment. Des logiciels tels que **T**ableau **Im**ager ou TIM ainsi que la suite Symantec Ghost. Pour le bien du projet, j'ai aussi téléchargé différents logiciels tels qu'*Autopsy,* un logiciel permettant l'analyse. Ces logiciels de même que les tests réalisés seront présentés plus en détails dans les parties suivantes.

2.1.2 Tableau Imager, TIM

Comme dit auparavant, la machine FRED était fournie avec un logiciel de création appelé **T**ableau **IM**ager. Ce logiciel a été développé par l'entreprise Tableau qui a ensuite été rachetée par Open Text, il permet de créer des images de support pouvant contenir des données.

Imager un disque dur permet de le cloner, d'en créer une copie plus légère et sous différents formats permettant de la stocker facilement. Cette image peut permettre en cas de problèmes liés au système d'exploitation de restaurer une version précédente afin de ne pas tout perdre. C'est très utile en cas d'attaque par ransomware, attaque ayant généralement pour but de chiffrer les données et de réclamer une rançon ensuite.

Pour fonctionner, ce logiciel a besoin d'une baie telle que celle présente sur la machine FRED qui permet de bloquer l'écriture sur un disque dur et ainsi de ne pas corrompre ou d'altérer les données. Le logiciel peut aussi fonctionner avec le matériel vendu par l'entreprise Open Text, matériel ayant le même principe de fonctionnement que la baie 4D.

La procédure pour créer une image est assez simple. Il suffit de brancher le disque ou le support contenant des données à analyser à la baie 4D en utilisant les différents adaptateurs et câbles mis à disposition. Après avoir correctement branché le support, il suffit de lancer le logiciel. Le disque apparaîtra et sera disponible pour être imagé

La création d'image peut avoir différentes utilités telles que l'analyse forensique ou la sauvegarde d'un disque dur afin de le recréer sur un autre support.

2.1.3 Autopsy, Digital Forensics

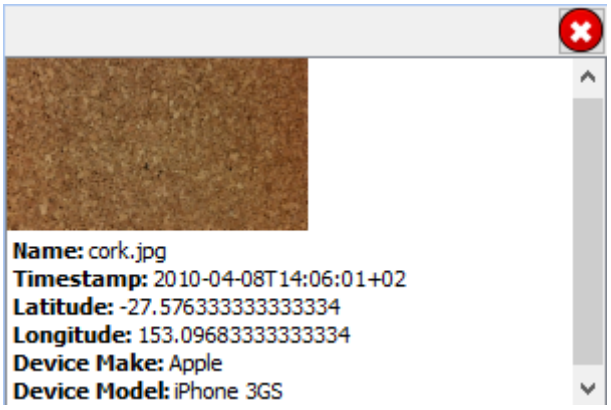
Autopsy est un logiciel de forensique open-source basé sur la suite *Sleuth Kit*, suite Linux permettant l'analyse de disques durs, développé par Basis Technology et en grande partie par Bryan Carrier. C'est un logiciel permettant l'analyse en profondeur d'une image disque en utilisant différents outils.

Les différents outils présents sur le logiciel sont :

- De la géolocalisation, en utilisant les métadonnées des différentes photographies qui ont pu être prises et stockées sur le disque.
- Un outil de visualisation des communications par mails, en utilisant les différentes adresses mails et les différents messages ayant transité par le conteneur.
- Un graphique temporel présentant les différentes années et le nombre de fichiers créés durant ces années avec des options plus détaillées et le moyen de réduire l'échelle temporelle.
- Un outil permettant la recherche de vidéo et d'image en utilisant différents filtres.

- Une arborescence avec différents filtres permettant d'afficher soit le chemin vers le fichier, soit le trier et le regrouper avec d'autres fichiers sous le même format.

Grâce à Autopsy, la récupération de fichiers supprimés est possible. Il suffit de trouver le fichier dans l'arborescence et le fichier sera obtenable seulement s'il n'a pas été réécrit. À moins que ces fichiers aient été réécrits, ils seront toujours analysables et accessibles en utilisant des logiciels spécialisés.



Un exemple de géolocalisation avec des métadonnées

2.1.4 Inventaire

Afin de bien prendre en main et comprendre le fonctionnement de la machine FRED, j'ai réalisé un inventaire des différents outils, câbles et adaptateurs présents sur la machine et dans la boîte à outil fournie avec cette dernière.

Pour réaliser cet inventaire, je me suis servi de la boîte à outils contenant les câbles et adaptateurs ainsi que les différentes informations présentes sur le site officiel des fournisseurs *Digital Intelligence* et *Tableau*.

Grâce à cet inventaire, j'ai pu comprendre comment relier l'Ultra Bay 4D à différents types de disques durs ainsi que d'autres appareils pouvant contenir des données. Vous trouverez l'inventaire en annexe.

Durant cet inventaire, j'ai pu comprendre l'utilité de chacun des câbles et des différents appareils après une suite de recherches étant donné que je n'avais jamais vu de câbles aussi spécifiques.

Peu de temps après avoir fini l'inventaire, j'ai pu obtenir des disques durs venant de la direction informatique de l'ENSI, ce qui m'a permis de commencer la phase de tests et de création d'image.

2.1.5 Création d'image mémoires

Pour faire une analyse de disque dur, il faut obligatoirement un appareil permettant de bloquer l'écriture afin de ne pas fausser les données. C'est pourquoi cela ne fonctionne pas sur des appareils lambda si vous essayez d'utiliser le même logiciel que j'ai utilisé.

Dans le lot de disques durs donnés par la DSI, j'ai décidé d'en sélectionner 4, chacun de marques différentes, et de les faire imager en les branchant à la baie. Sans documentation ni procédure pour réaliser cette tâche, j'ai dû tout réaliser en autonomie et découvrir le logiciel par

moi-même, ce fut assez simple ; en effet, le logiciel est simple d'utilisation et à comprendre globalement.

Pour relier la baie et les disques durs, j'ai dû utiliser un câble Tableau TC4-8-R3 avec un double câble SATA/SAS unifié ainsi qu'un connecteur et un port d'alimentation. Ensuite, il suffit de lancer le logiciel fourni avec la baie que j'ai détaillé plus haut, **TIM ou Tableau IMager**. Une fois le logiciel lancé, l'acquisition du disque dur est faisable. Chacune de ces acquisitions m'a pris un peu de temps, car les disques durs étaient assez volumineux. Cette manipulation a été facilitée grâce à l'inventaire que j'avais fait auparavant. En effet, ce dernier m'a permis de comprendre quels câbles utiliser pour relier différents composants à la baie. Les résultats de ces analyses étaient aussi volumineux, j'ai donc dû les stocker dans un des disques durs présents dans les racks.

Le logiciel Tableau Imager possède d'autres fonctionnalités telles que la création de rapport d'informations sur un disque dur.

Suite à des réunions, nous avons fixé d'autres objectifs et missions. L'objectif d'acheter des disques durs à une boutique de revente et de les analyser dans le cadre de test, différents de ceux effectués avec les disques durs de l'ENSI, un autre projet était de créer un script rapide permettant de transformer les rapports de Tableau Imager qui étaient sous format Log ou texte en format YAML.

2.1.6 Automatisation des rapports

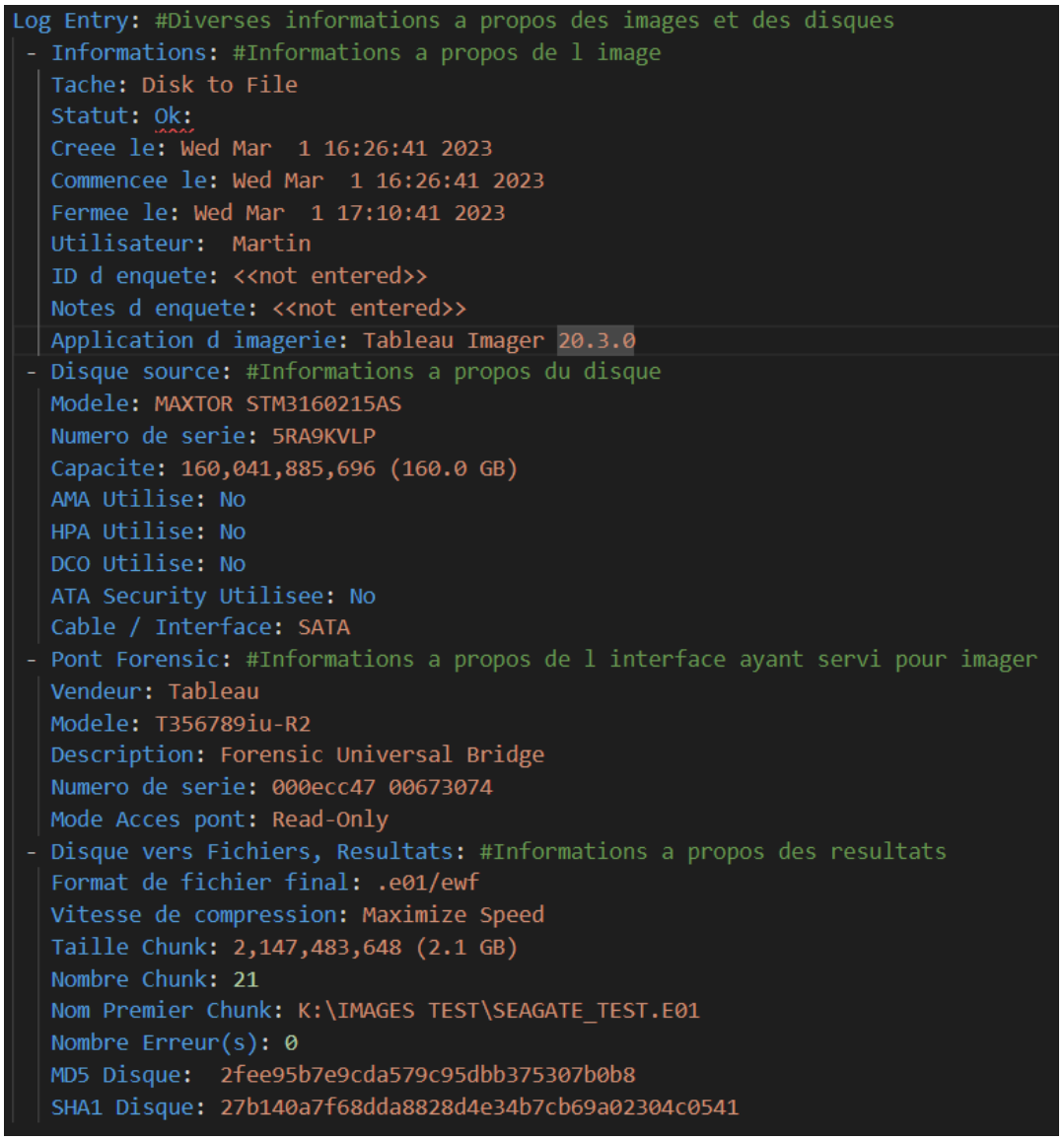
Ce fût une tâche intéressante, n'ayant pas fait de programmation depuis un certain temps, elle me permit d'entretenir mes connaissances. Pour la réalisation du script, j'utilisais Python, un langage simple et basique. Le script fut simple à réaliser malgré quelques détails sur lesquels j'ai pu bloquer, notamment l'utilisation de **Reg**ular **Ex**pression, une notion permettant la sélection de morceaux de texte d'après un schéma précis.

Le script était basique, une lecture d'un document et la suppression des informations et des lignes qui n'étaient pas nécessaires puis la réécriture dans un format différent.

Le format de base du fichier facilitait la création du script. En effet, chacune des informations étant écrites sur une ligne, l'ouverture et la lecture du fichier par des commandes python était simplifiée, je pus alors créer une liste dont chaque élément contient une ligne du fichier originel. Les éléments qui n'étaient pas nécessaires sont supprimés. Les lignes inutiles sont supprimées par une simple suppression d'éléments. Le reste est supprimé en passant par deux filtres. La réécriture du document se fait simplement par ouvrir un fichier et l'écriture des informations dans ce document. Le document sous format YAML est ensuite déplacé vers un autre dossier.

J'ai dû faire évoluer le script de nombreuses fois, en effet, j'ai pu rencontrer différents problèmes à cause de la méthode que j'ai pu utiliser. Par exemple, mon premier essai ne comprenait pas de description ni de numéro de "Case", lors d'un second test, j'ai réalisé l'erreur et fit évoluer le script avec une méthode qui permettait de déterminer.

Ces rapports ayant pour objectifs de référencer chacune des modifications et des manipulations apportées, ce qui est nécessaire dans le cadre d'enquêtes de police ou bien même d'archivage de documents. Ce sera plus simple à exploiter et à intégrer au site internet sous le format YAML, c'est pour cela que j'ai choisi ce format.



```
Log Entry: #Diverses informations a propos des images et des disques
 - Informations: #Informations a propos de l image
   Tache: Disk to File
   Statut: Ok:
   Creee le: Wed Mar  1 16:26:41 2023
   Commencee le: Wed Mar  1 16:26:41 2023
   Fermee le: Wed Mar  1 17:10:41 2023
   Utilisateur:  Martin
   ID d enquete: <<not entered>>
   Notes d enquete: <<not entered>>
   Application d imagerie: Tableau Imager 20.3.0
 - Disque source: #Informations a propos du disque
   Modele: MAXTOR STM3160215AS
   Numero de serie: 5RA9KVLP
   Capacite: 160,041,885,696 (160.0 GB)
   AMA Utilise: No
   HPA Utilise: No
   DCO Utilise: No
   ATA Security Utilisee: No
   Cable / Interface: SATA
 - Pont Forensic: #Informations a propos de l interface ayant servi pour imager
   Vendeur: Tableau
   Modele: T356789iu-R2
   Description: Forensic Universal Bridge
   Numero de serie: 000ecc47 00673074
   Mode Acces pont: Read-Only
 - Disque vers Fichiers, Resultats: #Informations a propos des resultats
   Format de fichier final: .e01/ewf
   Vitesse de compression: Maximize Speed
   Taille Chunk: 2,147,483,648 (2.1 GB)
   Nombre Chunk: 21
   Nom Premier Chunk: K:\IMAGES TEST\SEAGATE_TEST.E01
   Nombre Erreur(s): 0
   MD5 Disque:  2fee95b7e9cda579c95dbb375307b0b8
   SHA1 Disque: 27b140a7f68dda8828d4e34b7cb69a02304c0541
```

Fichier YAML, résultant du script

Voici par exemple le résultat d'un rapport sous le format YAML. Vous verrez ci-dessous le format original, en LOG, incomplet, mais avec certaines informations. Avant d'avoir ce résultat, j'ai imaginé un format et les différentes informations qui sont nécessaires à ces rapports, le format fût rapidement créé et en voici le résultat.

```
----------------------Start of Tableau Imager Log entry-----------------

Task: Disk to File
Status: Ok
Created: Wed Mar  1 16:26:41 2023
Started: Wed Mar  1 16:26:41 2023
Closed: Wed Mar  1 17:10:41 2023
Elapsed: 44 min
User:
  Martin
Case ID: <<not entered>>
Case Notes: <<not entered>>

Imager App: Tableau Imager
Imager Ver: 20.3.0

-------------------------Source Disk---------------------------

Model: MAXTOR STM3160215AS
S/N: 5RA9KVLP
Firmware Revision: 4.AA
Capacity in bytes reported Pwr-ON: 160,041,885,696 (160.0 GB)
Capacity in bytes reported by AMA: 160,041,885,696 (160.0 GB)
Capacity in bytes reported by HPA: 160,041,885,696 (160.0 GB)
Capacity in bytes reported by DCO: 160,041,885,696 (160.0 GB)
AMA in use: No
HPA in use: No
DCO in use: No
ATA Security in use: No
Cable/Interface type: SATA
Blank check status: Not Blank
```

Fichier de description du disque dur et de la procédure utilisée

Le script permet de faciliter la lecture des données des rapports ainsi que la vérification de ces dernières dans le cadre de procédures judiciaires durant lesquelles toutes informations de date et d'intégrité doivent être vérifiables. Ces données seront aussi ensuite utilisées par la plateforme web.

2.1.7 Analyses des images mémoires

Après avoir fini les différentes acquisitions et en étant en possession d'images de disques durs à utiliser, je pus commencer la phase d'analyse d'image. Cette phase prit plus de temps que prévu suite aux différents problèmes que je rencontrai, en effet le logiciel prenait plusieurs heures à faire l'analyse de l'image et pouvait se couper durant les analyses, ce qui m'a forcé à recommencer plusieurs fois.

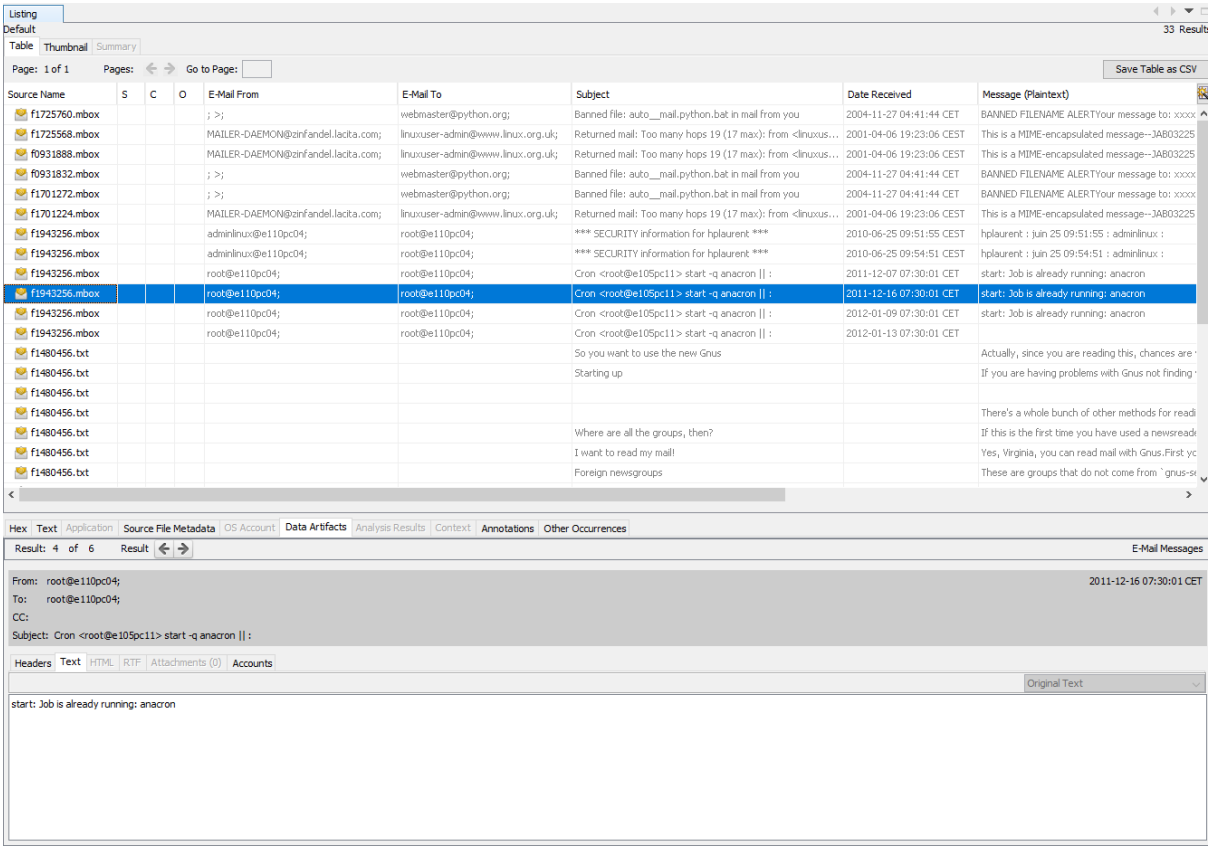
Analyser une image m'aura pris beaucoup de temps, plus d'une semaine pour être exact, j'ai dû stopper l'analyse avant la fin, car le logiciel utilisé n'avançait plus et ne répondait plus. Malgré tout, le résultat fut concluant et intéressant. Une vaste sélection de fichiers était présente sur le disque, ce qui m'a permis de faire de nombreux tests et d'utiliser les différents outils mis à disposition. Ces filtres formant une arborescence complète du système de fichier ainsi qu'une partie avec la possibilité de trier par format de fichiers.

Parmi ces différents outils, j'ai pu en utiliser une grande majorité. Ce qui m'a permis de comprendre comment fonctionnait globalement le logiciel ainsi que les métadonnées, utilisées pour la géolocalisation et d'autres systèmes de données que je ne connaissais pas, mais qui s'avéraient primordiales dans le domaine de la forensique.

L'outil de géolocalisation était l'un des plus intéressants à mes yeux même s'il n'y avait pas beaucoup de données à analyser. Le principe était assez simple sur cet outil, une grande carte s'affichait à l'écran et des points de différentes couleurs permettaient de localiser les différentes photographies. En cliquant sur chacun des points, la photographie apparaissait accompagnée de différentes informations telles que la description technique de l'appareil ayant pris ces photos.

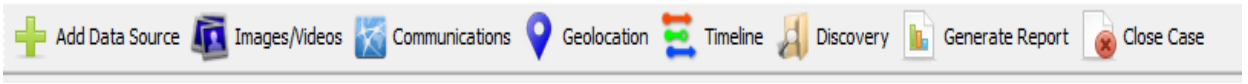
Il existe aussi un outil permettant la visualisation des communications entre mails ainsi qu'entre numéros de téléphones étant passés par le support analysé. Cet outil était, tout comme l'outil de géolocalisation, l'un des plus intéressants. Cet outil permet aussi de créer des profils de personnes liés aux

adresses mails, des informations annexes étaient aussi ajoutables. Après des recherches, j'ai pu découvrir qu'Autopsy se servait des traces de fichiers sous format MBOX, format utilisé principalement par les applications d'envois de mail, pour récupérer les mails et ainsi les visualiser. Comme nous avons pu le voir durant la matière sur le mail, le logiciel Thunderbird utilise ce format et laisse donc des traces analysables.



Interface de visualisation de mail
Retrouvable en annexe

D'autres outils étaient présents, mais moins intéressants, par exemple un outil d'histogrammes de fichier ou un outil de triage de vidéos et images.



Différents outils présents sur Autopsy

J’ai aussi utilisé différents modules externes téléchargeables ayant différentes utilités, tel que des plug-ins de reconnaissances faciales, qui n’ont malheureusement pas fonctionnées sur les images que j’utilisais, mais qui ont fonctionnées sur les sauvegardes de tests.

À la fin des différentes utilisations du logiciel, j’ai aussi testé de générer des rapports sous différents formats. Ces rapports résumant les différents types et formats de fichiers trouvés sur le disque. Ils étaient très complets et permettaient d'appréhender les résultats des enquêtes plus simplement qu’en ouvrant le logiciel.

J’ai trouvé ces manipulations et recherches très instructives, car la forensique était encore un nouveau sujet pour moi, n’en n’ayant jamais fait auparavant. J’ai aussi pu comprendre comment se déroulaient des enquêtes qui nécessitaient de la recherche d’informations sur des disques durs, et ce, via mes recherches sur le sujet.

Grâce à des recherches ayant pour but de comprendre le fonctionnement d’*Autopsy*, j’ai pu comprendre comment les fichiers supprimés sont récupérés. Le principe est assez simple, lorsqu’un fichier est “supprimé”, il ne l’est pas vraiment ; seulement son chemin pour l’ouvrir et le trouver l’est, ses données sont toujours “écrites”. Le seul moyen de le supprimer définitivement est de “réécrire” à blanc par-dessus ces données.

Durant ces recherches, j’ai aussi pu comprendre comment les communications par mail ainsi que les adresses mails utilisées étaient trouvées. Le module de visualisation cherche les traces de fichiers MBOX, EML et PST, des fichiers de mail et les regroupent pour les analyser.

2.1.8 Rédactions de méthodologie et de documentations

En parallèle des différentes manipulations que j'ai faites, j'ai aussi rédigé des documentations sur les logiciels utilisés ainsi que des méthodologies sur le processus des manipulations. Ces documentations et méthodologies serviront aux futures personnes souhaitant analyser des images et en créer.

Les différentes documentations sur les logiciels et sur la machine FRED furent les premières tâches que j'ai réalisées en attendant d'avoir les disques durs, ce qui m'a permis de comprendre le fonctionnement des logiciels, tandis que l'inventaire cité précédemment m'a permis de comprendre comment relier les supports et la machine.

Même sans avoir d'expérience dans la rédaction de méthodologie, je me suis inspiré des différents comptes-rendus ainsi que les sujets de TP que j'ai pu faire jusqu'à présent à l'IUT, ce qui m'a facilité le travail.

Ces méthodologies et procédures ont pu être testées dans la phase de test qui suivra.

2.1.9 Phases de test

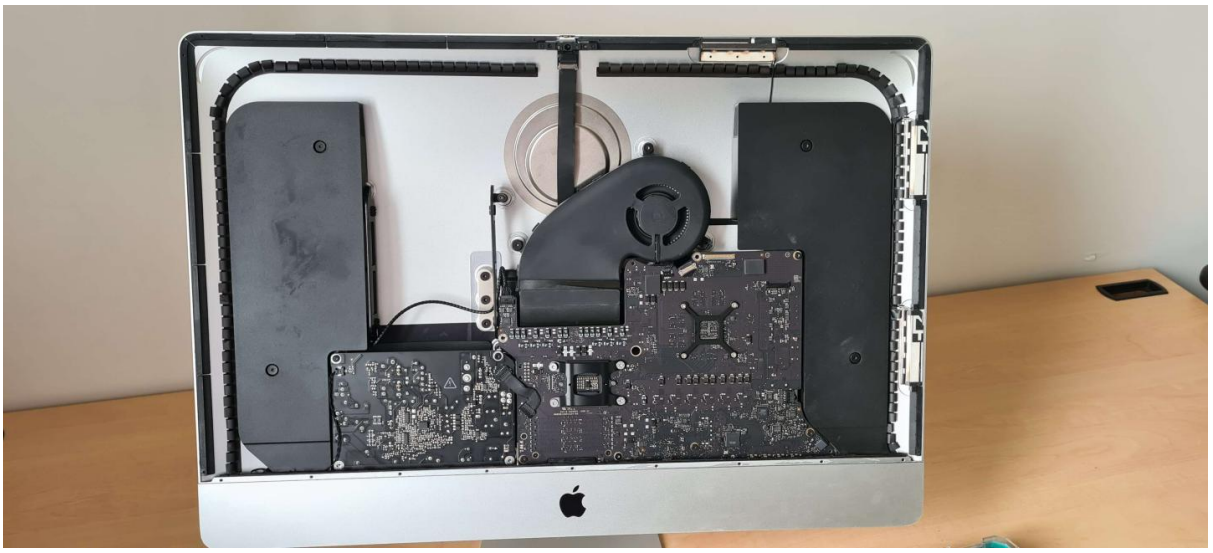
J'ai pu réaliser différentes phases de tests durant mon stage, certains consistaient en simples tests de l'interface web en tant qu'utilisateur lambda ainsi qu'en tant qu'administrateur. L'objectif de ces tests étaient de trouver des possibles bugs et aussi d'apporter un avis extérieur à l'interface et son design afin de les améliorer.

Les tests sur l'interface web ainsi que les filtres étaient assez simples, en premier lieu, je testais des fonctionnalités simples, le moindre bug ou problème était noté pour être ensuite réglé. Après quelques jours de tests simples, nous avons décidé de faire d'autres tests en cumulant plusieurs fonctions et filtres. Plusieurs bugs et problèmes ont pu être décelés durant ces tests, ce qui a permis d'améliorer l'interface web et les différents

programmes. En tant qu'utilisateur, j'ai aussi pu apporter mon avis sur le début d'interface et les différents graphiques présents dessus.

J'ai aussi demandé à un des ingénieurs de l'équipe de venir tester les procédures que j'avais rédigées et de me donner un retour pour les améliorer. Ce fut assez rapide, ces procédures étant courtes. Je pus ensuite corriger les erreurs présentes ainsi qu'améliorer en termes de clarté ces documents.

Durant ces tests, j'ai aussi pu démonter un iMac du laboratoire qui était inutilisé, l'objectif étant d'atteindre le disque dur à l'intérieur de cet ordinateur. J'ai ensuite imagé ce disque dur dans un format spécifique à Apple. Ne savant pas ou le disque était situé, j'ai démonté l'entièreté de l'iMac. J'ai beaucoup aimé cette tâche qui m'a permis de comprendre comment était fait un ordinateur iMac, fonctionnant et étant construit légèrement différemment des ordinateurs normaux.



Photographie MAC que j'ai pu démonter

2.1.10 Réunions de projet

En tant que membre du projet G'DIP, j'assistais aussi aux réunions qui avaient lieu environ toutes les deux semaines environ. Durant ces dernières, l'objectif était de fixer des tâches à faire entre chacune de ces réunions ainsi que de se concerter sur les différentes avancées. Ces réunions permettaient aussi de se concerter afin d'avancer mutuellement dans le projet et de trouver des axes d'amélioration.

Pour faire un parallèle, j'ai trouvé que ces réunions ressemblaient à celles que nous avions organisées durant le projet du premier semestre parmi le groupe Bravo. Elles servaient de compte rendu et de mise en commun du travail fait, ce que j'ai trouvé très intéressant.

2.1.11 Recherches liées à la Forensique

Durant les différentes parties du projet G'DIP détaillées ci-dessus, j'ai eu l'occasion de faire des recherches sur différents sujets, telles que les plug-ins autopsy, ou même des machines de Forensique telles que celles produites par l'entreprise Celebrite.

Ces recherches m'ont mené à découvrir qu'en termes de Forensique, Israël était l'un des pays les plus avancés. Ce sont eux qui ont produit les différentes machines fournies aux commissariats français ainsi qu'à d'autres institutions telles que le FBI ou bien même Interpol. Ces machines ayant comme utilisé la récupération de données de téléphones ou bien même d'ordinateurs. Cellebrite ont aussi développé différents logiciels sous licence facilitant l'analyse des données récupérées précédemment. Les logiciels et machines étant payants, je n'ai pas pu les tester pour essayer de comprendre leurs fonctionnements. J'ai aussi pu découvrir que l'entreprise fut mêlée à différentes affaires. La plus connue étant un conflit provoqué par le dirigeant de l'entreprise Signal, Moxie Marlinspike, qui a dévoilé plusieurs failles dans les logiciels de Cellebrite ainsi qu'une utilisation illégale de certains fichiers d'Apple.

Ces recherches ne furent qu'annexes, mais restèrent intéressantes afin de se renseigner sur les différents acteurs du monde de la forensique.

2.1.12 Vidéo Démonstration

Suite à des grèves, les journalistes censés venir au laboratoire pour interviewer les membres du projet G'DIP ont été forcés de se désister. Après en avoir parlé durant une réunion, nous avons convenu, monsieur Rosenberger et moi, qu'afin de permettre la présentation de la machine FRED à des intervenants externes, je devais faire une courte vidéo. J'ai donc décidé de filmer une démonstration de la machine, en branchant un disque dur et en lançant l'analyse, j'ai ensuite fait un montage rapide sur cette courte vidéo.

2.2 Léo Rover

Le Léo Rover est un petit robot créé et développé par l'entreprise Fiction Lab. La pièce centrale du robot étant une carte Raspberry pi, une sorte de petit ordinateur, fonctionnant sous Ubuntu. Ce fût simple de prendre en main l'appareil. D'autres pièces viennent s'ajouter, une caméra diffusant un flux vidéo en continu, un GPS d'intérieur ainsi que différents radars et détecteurs.

J'ai aussi pu découvrir que la carte Raspberry hébergeait un site Ethernet disponible sur son adresse IP qui permettait de contrôler le robot à partir de joystick et de se diriger à partir du rendu de la caméra. Sur ce même site, différentes informations à propos des capteurs et de la batterie du robot sont disponibles.

Le projet autour du robot était de le paramétrer pour qu'il se déplace automatiquement à travers l'étage dans lequel l'équipe SAFE évolue et pour qu'il réagisse avec un message personnalisé après avoir croisé une personne. Ce projet n'en était pas vraiment un, en effet, c'était plus une activité annexe permettant de m'occuper durant mon temps libre ou lors des analyses. Même si je n'ai pas pu le réaliser en entier, le travail fourni et les recherches faîtes étaient intéressantes et m'ont permis de m'améliorer dans le domaine de la programmation.

Après en avoir parlé avec mon maître de stage, il m'a expliqué que les différents projets sur le rover était de le faire se déplacer automatiquement vers différents lieux sur lesquels se déroulent des groupements à partir de caméra connectée qui détectent la présence de personnes et enverraient une commande au robot lui signalant de venir sur place, un autre projet étant de permettre au robot de guider les personnes dans le laboratoire en associant des coordonnées sur un plan à des lieux du laboratoire.

2.2.1 Prise en main du robot

Pour commencer à prendre en main le robot, j'ai décidé en premier lieu de faire des recherches, ces dernières m'ont menée à trouver un site internet avec différents modes d'emplois. J'ai commencé par faire les tâches les plus simples du robot, c'est-à-dire le faire avancer et reculer en ligne de commande. Ensuite, j'ai accentué la difficulté en écrivant un code qui lui permettait d'avancer, reculer et tourner en exécutant des commandes.

Ces deux tâches que je me suis fixées m'ont permis de bien comprendre le fonctionnement du Rover et m'ont aussi permis de planifier et de créer un schéma sur ce que j'allais faire pour le reste du projet.

Sur le site mentionné ci-dessus, j'ai pu trouver une méthode pour que la conduite se fasse de manière semi-automatique, en effet, il fallait que le rover repère son environnement pour se déplacer après. Pour ce faire, il suffisait d'installer différents paquets et de se connecter en SSH pour ensuite configurer le rover et paramétrer les différents paquets et pièces ajoutables sur le rover par exemple le radar Li Dar.

2.2.2 Tâche principale

N'étant pas une tâche principale, je n'y ai accordé que peu de temps, mais malgré tout, j'ai pu produire un schéma ainsi que commencer un script sur la conduite automatique, tout en développant des idées. L'idée était de faire tourner le rover dans tout l'étage et dès qu'un obstacle était rencontré, faire une salutation personnalisée si le visage était reconnu sinon faire dire "bonjour" et demander si la personne avait besoin d'aide ou tout simplement esquiver l'obstacle en se décalant.

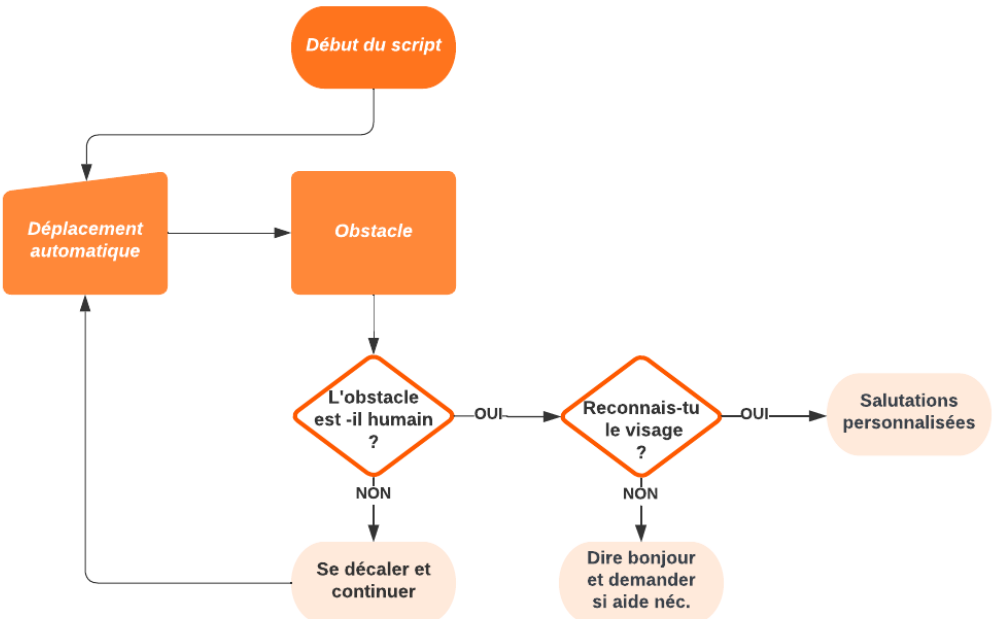


Schéma du projet, fait moi-même

N'ayant jamais de module de reconnaissance faciale, j'ai dû en chercher un et l'utiliser en le liant avec une base de données contenant les différents visages et noms. J'ai donc décidé d'utiliser le module "facial-recognition" qui permet de détecter les différents points d'un visage sur une photographie ainsi que de lier ce visage à un nom ou une personne. Sachant qu'il y a une caméra sur le rover et que cette dernière renvoie le flux vidéo sur ce qui est appelé un "topic", il me suffirait de lancer l'analyse de reconnaissance faciale durant la rencontre d'un obstacle.

Le robot ayant déjà été utilisé, j'ai pensé me servir d'un script déjà présent pour la reconnaissance de voix ; je devais seulement trouver le moyen de l'ajouter à la partie principale. Il ne me restait donc que la détection d'obstacles que je ne savais pas faire malgré ces difficultés, j'ai essayé de chercher des pistes, malheureusement.

2.3 Séminaires et réunions

2.3.1 Séminaires

Durant ma présence au sein de l'équipe SAFE, j'ai pu assister aux différents séminaires informels organisés pour l'équipe. Ces séminaires se déroulaient lors de la pause du midi et l'entièreté des membres étaient conviés. Différents sujets ont pu être abordés, chacun étant différent du précédent. Les intervenants étaient pour la plupart sélectionnés lors des sessions précédentes. J'ai moi aussi participé en tant qu'orateur à l'un de ces séminaires ; j'ai pu présenter mon avancée sur la machine FRED et la compréhension de son fonctionnement ainsi que l'entreprise Cellebrite et les différentes affaires dans lesquelles cette entreprise fut mêlée.

J'ai aussi pu assister à différents séminaires organisés pour l'équipe SAFE avec différents intervenants venus présenter leurs thèses. Différents sujets ont pu être abordés et même sans avoir des connaissances précises, ces présentations étaient intéressantes et m'ont permis de découvrir d'autres domaines et sujets que je ne connaissais pas.

2.3.2 Réunions

En plus des séminaires, j'ai pu assister à différentes réunions, deux pour être exact. La première avait pour but de faire un point au niveau de l'équipe, des différentes thèses et des projets en cours et à venir, du budget ainsi que de la politique de recrutement. N'ayant jamais assisté à de réunions d'administration telles que celle-ci,

3 - Conclusion des projets

3.1 Projet G'DIP

Le projet G'DIP fut très intéressant et me permit de m'introduire au monde de la forensique, une discipline que je ne connaissais que de nom. J'ai pu comprendre ce qui composait ce domaine, sa part importante dans tous les domaines de la justice ainsi que de la conservation de documents.
Le projet continuera, même après mon départ. J'espère avoir pu apporter ma pierre à l'édifice avec ma participation. J'espère aussi que les différents documents et tests que j'ai pu réaliser vont être utiles pour les différentes personnes qui utiliseront la machine.

3.2 Léo Rover

Le projet sur le Rover n'en était pas vraiment un, c'était principalement une activité annexe, mais j'ai tout de même trouvé cela intéressant et ce projet m'a permis d'utiliser mes connaissances acquises dans la matière du troisième semestre R310 "Internet Of Things", ainsi que des connaissances de programmations. J'espère aussi que mes différentes recherches et mes programmes seront utiles pour le futur.

4 - Bilan du stage

Durant ces dix semaines de stage, j'ai pu découvrir la vie en laboratoire, différente de celle en entreprise, au sein du GREYC. Grâce à ce stage, j'ai aussi pu découvrir différents postes et voies futures que je ne connaissais pas auparavant. Ce stage fût une ouverture vers un monde qui m'était inconnu jusqu'à maintenant.

N'ayant jamais abordé la forensique comme thème de la cybersécurité, je trouve que ce stage fût une très bonne introduction à ce domaine dont les racines s'ancrent comme la base de la cybersécurité. Étudier la machine FRED et faire des procédures pour aider les prochains utilisateurs de la machine furent des tâches très intéressantes et me permirent de développer de nouvelles méthodes de travail en autonomie ainsi qu'en coopération avec des collègues.

Ce stage me permit de discuter avec des professionnels du milieu de la cybersécurité et de la recherche en informatique, ce qui me conforta dans mon choix d'études tout en éclaircissant les différents horizons qui se présentent à moi.

Ce stage m'aura beaucoup apporté et je remercie encore une fois Christophe Rosenberger de m'avoir accueilli au sein du laboratoire. Ce fut une très bonne expérience professionnelle qui m'aura beaucoup appris sur le monde de la recherche en laboratoire.

J'espère que mon stage au laboratoire GREYC ouvrira une porte aux futurs étudiants et qu'ils feront aussi des stages dans des laboratoires liés à l'ENSI ou à l'Université de Caen.

5 - Glossaire

- Forensique : Investigation d'un système d'information, analyse de supports numériques
- Racks : Outil de rangement permettant de stocker des disques durs et autres appareils de stockage de données.
- FRED : Forensique Recovery of Evidence Device
- YAML : Yet Another Markup Language ou YAML Ain't Markup Language, langage de représentation de données.
- Topic : terme utilisé en IoT (Internet Of Things), définissant un canal de parution de données.
- MBOX / PST / EML : Formats de stockage de mail

6 - Bibliographie & webographie

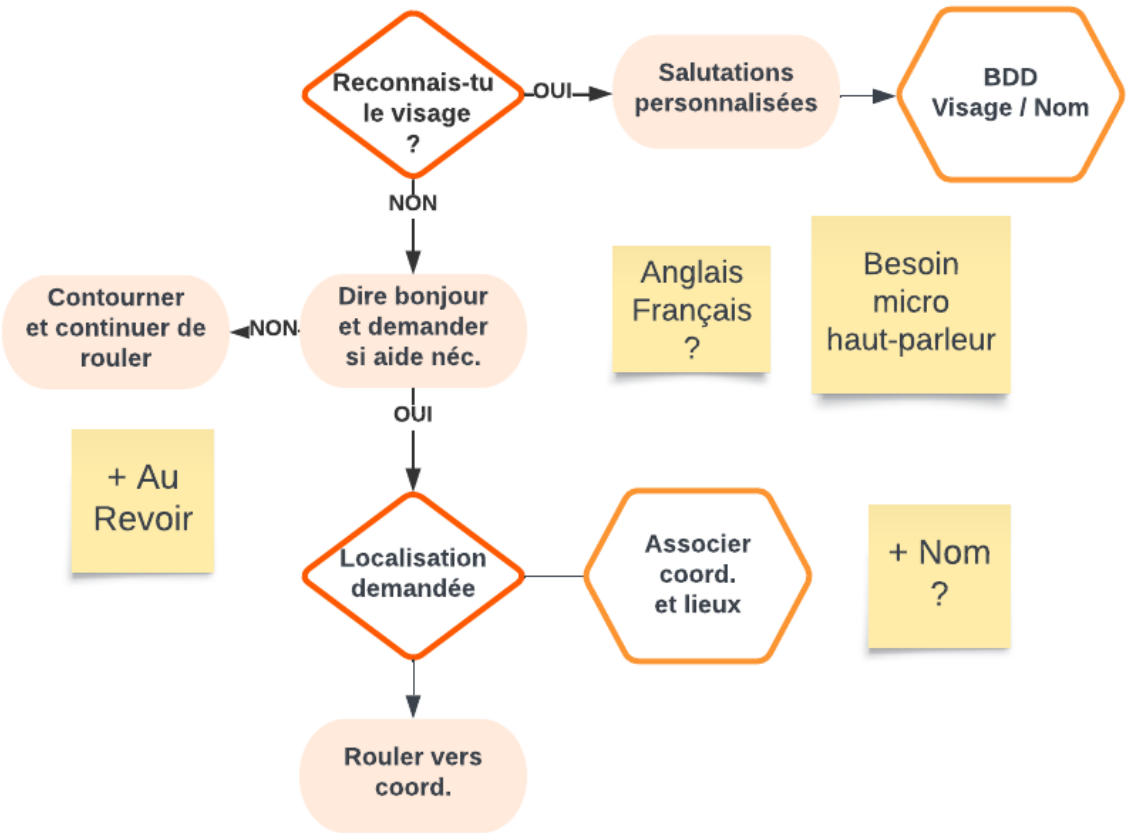
- FRED : https://digitalintelligence.com/products/fred/
- GREYC : https://www.greyc.fr/
- Différents câbles inventaire : https://security.opentext.com/tableau/hardware/
- Autopsy : https://www.autopsy.com/ & https://www.sleuthkit.org/
- MBOX : https://fr.wikipedia.org/wiki/Mbox
- PST : https://fr.wikipedia.org/wiki/.pst
- EML : https://www.malekal.com/qu-est-qu-un-fichier-eml-comment-ouvrir/
- Léo Rover : https://www.leorover.tech/
- face-recognition : https://pypi.org/project/face-recognition/

7 - Annexes

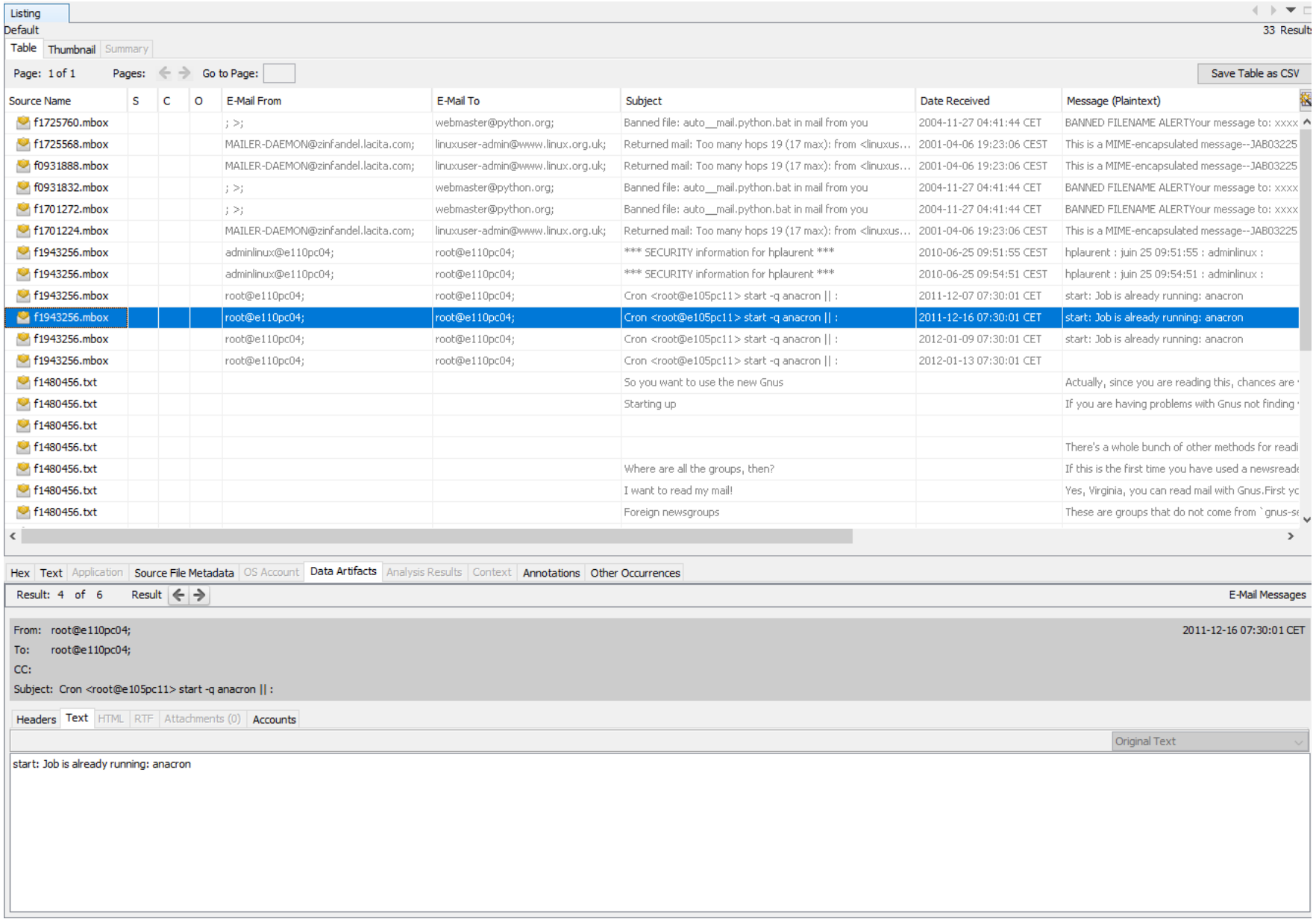
Inventaire cité dans la première partie :

		Nom	Quantité	Description
Boîte à outils	Câbles	TC2-8-R2	1	Câble avec un connecteur femelle à 4 broche de type "Molex" et un connecteur mâle 3M
		TC3-8	2	Câble SATA standard
		TC44-R3	1	Câble de signal avec un signal SATA/SAS unifié et un d'alimentation pour un SATA/SAS ainsi qu'un connecteur 3M d'alimentation
		TC54-R2	1	Câble avec un connecteur à 15 broches SATA d'alimentation et connecteur 3M
		TC6-8	1	Câble de haute qualité composé de 80 câble conducteur IDE et des connecteurs à 40 broches IDE
		TC7-9-9	1	Câble IEEE1394, FireWire800 avec 2 connecteurs 9 broches
		TC-PCIE-8	1	Câble fait pour connecter un adaptateur PCIE Tableau à un produit compatible
		USB3-AB-3	1	Câble avec un connecteur USB3 et un connecteur BM
	Adaptateurs	TDA1-1	1	Adaptateur pour une carte SSD
		TDA1-2	1	Adaptateur pour un SSD M2
		TDA1-3	1	Adaptateur pour un SSD Apple
		SATA IMICRO SATA	1	Adaptateur SATA vers MICRO SATA
		Bladetype SSD Adaptater	1	Adaptateur Mac Air BLADE type SSD vers SATA
		SATAIII toM2(NGFF) SSD	1	Adaptateur microSATA ou M.2 (SATA) vers SATA
		mSATA SSD Adaptater	1	//
		SATA LIF	1	Adaptateur SATA/LIF disque dur vers une connexion SATA
	Pièces de rechange	SLACK SATA	1	SLACK SATA semblable à ceux déjà présent dans la machine
		Mini-Ventilateur	1	Mini ventilateur ajoutable dans la machine
		Vis	4	Sachets de vis de différentes tailles à utiliser dans la machine
		Clefs	1	Clefs permettant d'ouvrir la machine
		Câble d'alimentation	1	Câble d'alimentation en double
		ASUS FAN HOLDER	1	Support servant pour le mini-ventilateur
		ASUS PIN	1	Câble de liaison
		SUPPORT	1	Pièce inconnue
FRED	Baie HOTSWAP	SATA	3	Slack enlevable de la machine à chaud permettant d'installer un disque dur / un SSD ect
		SLACK MICRO SD	1	Slack avec des slots de Micro SD, Compact Flash, SD/MMC, Memory Stick. Peut supporter des CFC, MSC, SMC, MD, xD, MSP, MPSD, SDC/SDHC/SDXC, MMC
	Baie SATA Drive	SATA	2	Slack enlevable à froid pareil à ceux cités plus haut
		IMAGING SHELL	1	Slack permettant de ventiler l'ordinateur
		UltaBay 4D	1	écran tactile et ports USB3, FIREWIRE, SATA, IDE, SAS
		BluRay BD-RE DVD+RW	1	Slack permettant d'insérer un disque bluray

Diagramme théorique plus poussé sur la reconnaissance vocale et les interactions avec le rover :



Machine FRED et installation utilisée durant mon stage :


Listing
Default
33 Results
Table
Thumbnail
Summary
Page: 1 of 1
Pages:
Go to Page:
Save Table as CSV
Source Name
S
C
O
E-Mail From
E-Mail To
Subject
Date Received
Message (Plaintext)
f1725760.mbox ; >; webmaster@python.org; Banned file: auto__mail.python.bat in mail from you 2004-11-27 04:41:44 CET BANNED FILENAME ALERTYour message to: xxxx
f1725568.mbox MAILER-DAEMON@zinfandel.lacita.com; linuxuser-admin@www.linux.org.uk; Returned mail: Too many hops 19 (17 max): from <linuxus... 2001-04-06 19:23:06 CEST This is a MIME-encapsulated message--JAB03225
f0931888.mbox MAILER-DAEMON@zinfandel.lacita.com; linuxuser-admin@www.linux.org.uk; Returned mail: Too many hops 19 (17 max): from <linuxus... 2001-04-06 19:23:06 CEST This is a MIME-encapsulated message--JAB03225
f0931832.mbox ; >; webmaster@python.org; Banned file: auto__mail.python.bat in mail from you 2004-11-27 04:41:44 CET BANNED FILENAME ALERTYour message to: xxxx
f1701272.mbox ; >; webmaster@python.org; Banned file: auto__mail.python.bat in mail from you 2004-11-27 04:41:44 CET BANNED FILENAME ALERTYour message to: xxxx
f1701224.mbox MAILER-DAEMON@zinfandel.lacita.com; linuxuser-admin@www.linux.org.uk; Returned mail: Too many hops 19 (17 max): from <linuxus... 2001-04-06 19:23:06 CEST This is a MIME-encapsulated message--JAB03225
f1943256.mbox adminlinux@e110pc04; root@e110pc04; *** SECURITY information for hplaurent *** 2010-06-25 09:51:55 CEST hplaurent : juin 25 09:51:55 : adminlinux :
f1943256.mbox adminlinux@e110pc04; root@e110pc04; *** SECURITY information for hplaurent *** 2010-06-25 09:54:51 CEST hplaurent : juin 25 09:54:51 : adminlinux :
f1943256.mbox root@e110pc04; root@e110pc04; Cron <root@e105pc11> start -q anacron || : 2011-12-07 07:30:01 CET start: Job is already running: anacron
f1943256.mbox root@e110pc04; root@e110pc04; Cron <root@e105pc11> start -q anacron || : 2011-12-16 07:30:01 CET start: Job is already running: anacron
f1943256.mbox root@e110pc04; root@e110pc04; Cron <root@e105pc11> start -q anacron || : 2012-01-09 07:30:01 CET start: Job is already running: anacron
f1943256.mbox root@e110pc04; root@e110pc04; Cron <root@e105pc11> start -q anacron || : 2012-01-13 07:30:01 CET
f1480456.txt So you want to use the new Gnus Actually, since you are reading this, chances are
f1480456.txt Starting up If you are having problems with Gnus not finding
f1480456.txt
f1480456.txt There's a whole bunch of other methods for readi
f1480456.txt Where are all the groups, then? If this is the first time you have used a newsreade
f1480456.txt I want to read my mail! Yes, Virginia, you can read mail with Gnus.First yc
f1480456.txt Foreign newsgroups These are groups that do not come from `gnus-se
Hex
Text
Application
Source File Metadata
OS Account
Data Artifacts
Analysis Results
Context
Annotations
Other Occurrences
Result: 4 of 6
Result
E-Mail Messages
From: root@e110pc04;
2011-12-16 07:30:01 CET
To: root@e110pc04;
CC:
Subject: Cron <root@e105pc11> start -q anacron || :
Headers
Text
HTML
RTF
Attachments (0)
Accounts
Original Text
start: Job is already running: anacron

8 - Résumés & mots-clés

Au cours de ma deuxième année de BUT, j'ai réalisé un stage au sein du laboratoire GREYC. Travaillant sur différents projets, notamment le projet G'DIP, projet ayant pour but de créer une plateforme pouvant servir aux enquêteurs faisant de la forensique. Travaillant avec trois autres personnes sur le projet, chacun ayant sa partie distincte. Ma partie étant la rédaction de documentations ainsi que de méthodologies, faire différents tests et la compréhension de la machine FRED.

En plus du travail fourni sur les différents projets, j'ai pu découvrir la vie d'un laboratoire, son organisation au quotidien ainsi que la gestion de projet.

J'ai beaucoup aimé ce stage qui m'a permis de comprendre le métier de chercheur et de m'introduire au milieu de la forensique.

During my second year of Bachelor, I did an internship within the GREYC Laboratory. Working on different projects, particularly the G'DIP Project, with the goal of developing a platform that would be used for investigators doing computer forensic. Working with three other people on the project, each one his own part. My part being writing documentation as well as methodologies, doing different tests and understanding FRED Machine.

In addition to the work on the various projects, I was able to discover the life within a laboratory, his daily organisation as well as project management.

I really liked this internship, which allowed me to understand the work of a researcher and introducing myself to the forensic world.

Mots-Clés : Forensic, Recherche, Laboratoire, Équipe, FRED