Pierre Mourlanne
Promotion 2014

Rapport de stages

Année de césure

Du 1 Août 2012 au 31 Janvier 2013 et du 12 Février 2013 au 9 Août 2013

Tables des matières

Tables des matières

Introduction : pourquoi la césure ?

Depuis la promotion 2013, tous les étudiants de SUPÉLEC doivent passer un semestre à l'étranger pour obtenir leur diplôme d'ingénieur. Pour répondre à cette exigence, nous pouvons aller dans une université partenaire, pour un semestre d'échange ou un double diplôme, ou bien faire un stage à l'étranger.

Les cours « Systèmes d'Information », « Architecture des Systèmes Informatiques » et le projet de fin d'année m'ont convaincu de rester à SUPÉLEC en deuxième année. La majeure « Systèmes d'Information Sécurisés », elle, m'a convaincu de rester pour ma dernière année. C'est pourquoi j'ai choisi de partir à l'étranger en stage plutôt que dans un cadre scolaire.

Par ailleurs, ayant beaucoup de mal à croire que les sept mois de cours en troisième année allaient faire de nous des experts en sécurité informatique, il me semblait indispensable de créer mes propres expériences avant de quitter l'école. C'est pourquoi j'ai choisi de faire cette année de césure.

Recherche et obtention des stages

Pour trouver mon stage à l'étranger, j'ai profité des contacts de Ludovic MÉ, que je ne remercierai jamais assez. Après l'avoir contacté, Christopher KRUEGEL m'a proposé un stage de six mois dans le laboratoire de sécurité informatique de l'université de SANTA BARBARA. Je n'ai pas eu à passer d'entretien, si ce n'est à l'ambassade américaine, où on s'est assuré que je n'étais pas un terroriste dealer de drogues (et de camembert).

Les recherches pour mon deuxième stage ont commencé lors de la fin de mon premier stage, depuis les États-Unis. Mon premier objectif était d'aller à l'ANSSI, pour continuer à travailler dans le domaine de la sécurité informatique. Mon profil de stagiaire en année de césure a peu intéressé l'ANSSI, qui préfère embaucher des étudiants dans le cadre de leur stage de fin d'études. J'ai alors répondu à une annonce sur l'Intranet de SUPÉLEC, postée par Sketchfab, qui cherchait un développeur Python/Django et Javascript. Je venais de faire du Python pendant 6 mois, j'avais pris le temps de suivre quelques cours de Javascript, j'ai donc contacté Pierre-Antoine Passet, ancien Supélec qui avait posté l'annonce.

J'ai ensuite passé un entretien à distance avec ce même Pierre-Antoine, directeur technique, puis avec Cédric Pinson, directeur général. Dans chacun de ces entretiens, nous avions discuté de mes compétences et motivations et j'ai pu poser mes questions sur le fonctionnement de Skechtfab. Une fois rentré en France début février, je suis allé les rencontrer au « Chaudron », une pépinière parisienne dans le dixième arrondissement. J'ai alors discuté avec Pierre-Antoine et Alban Desnoyel, directeur commercial. Nous avons parlé des modalités de mon éventuel stage suite à quoi j'ai accepté leur offre. C'était un jeudi, je commençais le lundi suivant.

Description factuelle : Seclab

Mon stage au laboratoire de sécurité informatique (Seclab) de l'université de Santa Barbara commençait le 1er août. Je suis arrivé aux États-Unis quelques jours plus tôt, ce qui m'a permis de m'habituer au changement de température, de fuseau horaire et d'échelle.

Mon premier jour était un mercredi, Christopher m'avait demandé de venir pour la réunion hebdomadaire, à 14h. Comme cela allait être le cas pendant les six mois suivants, toutes les personnes travaillant dans le laboratoire ont partagé un résumé de ce qu'elles avaient fait pendant la semaine. Plusieurs doctorants et « post-doc » revenaient de la DEF CON. Suite à cette courte réunion, une petite heure, des chercheurs du laboratoire m'ont présenté leurs projets de recherche. J'ai pu choisir parmi ces sujets. J'aurais également pu proposer une idée de projet. Étant plus confiant dans mes connaissances théoriques que pratiques, j'ai évité les sujets trop techniques (rétroingénierie par exemple). J'ai choisi de travailler avec Gianluca Stringhini et Manuel Egele sur la détection des spammers sur les réseaux sociaux dès l'étape d'identification.

On m'a ensuite aidé à m'installer dans le laboratoire, sur un bureau personnel, avec un ordinateur et un écran fournis par l'université. Manuel et Gianluca étant peu présents dans le laboratoire au début de mon stage, j'ai été rapidement lâché dans le grand bain.

Le début de mon travail a consisté en une lecture attentive des travaux de recherche sur la détection de spammers sur les réseaux sociaux. J'ai découvert que le sujet intéresse autant les chercheurs que les industriels. Ces travaux ont en particulier montré que les méthodes de détection actuelles sont facilement mises en défaut. En pratique, les spammers sont plus souvent repérés par les utilisateurs « légitimes » que par des méthodes automatiques. Ces utilisateurs malveillants ont donc souvent le temps d'agir avant d'être détectés. D'où la motivation de les détecter dès l'étape d'identification, avant qu'ils ne puissent mener leur attaque.

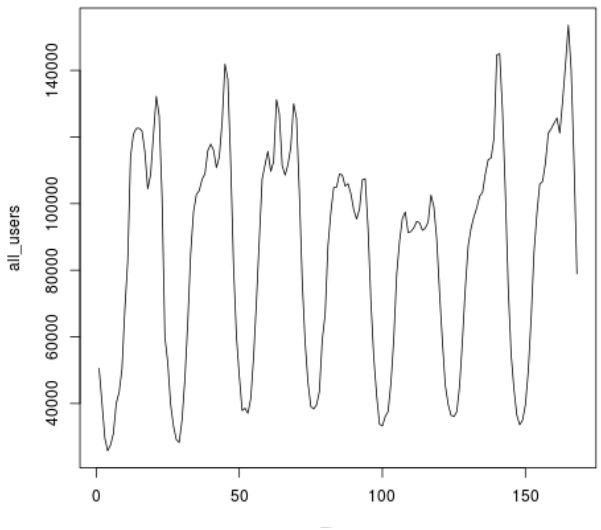
Je me suis ensuite familiarisé avec les données, fournies par Impermium. Ces données contenaient une semaine de logs d'identification. Chaque ligne était constituée d'une heure (jusqu'à la seconde), d'un identifiant chiffré, d'une adresse IP chiffrée et de l'en-tête HTTP (contenant notamment le « user-agent »). Gianluca avait déjà écrit un outil de décomposition analytique (ou « parser »), en Python. J'ai donc passé une petite semaine à apprendre le Python, notamment grâce aux tutoriaux de Zed Shaw.

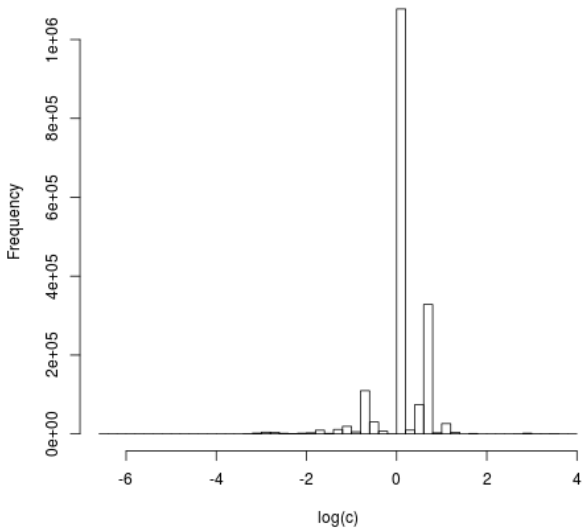
Habituellement, les études de comportements sur les réseaux sont menées sur un individu donné. Connaissant toutes ses actions sur le réseau social (ce que nous appellerons ses « données d'activité »), il est possible de créer un profil pour l'utilisateur. On étudie ensuite ce profil pour déterminer si cet individu est malveillant ou non. Ici, ne disposant que des données de connexion, il n'est pas possible de créer un profil suffisament complet pour chaque utilisateur. C'est pour cela que j'ai cherché à aggréger des utilisateurs, afin de disposer de suffisament de données pour créer un profil pour la communauté. Il me fallait encore trouver un critère pour former ces groupes d'individus.

J'ai choisi de regrouper les comptes en fonction du nombre d'adresses IP qu'ils partageaient. Mon hypothèse étant que si deux comptes différents sont, par exemple, accédés par une même adresse IP, il est probable qu'ils fassent partie d'un même réseau malveillant. En utilisant la merveilleuse librairie networkx, le travail de Thomas AYNAUD et des dizaines de Go de RAM, j'ai pu former ces communautés.

Dans un deuxième temps, il m'a fallu créer un profil pour ces agrégats d'utilisateurs. J'ai suivi deux axes de travail. D'une part, l'analyse temporelle et d'autre part l'étude de la corrélation entre « user-agent » et adresse IP. J'ai d'abord postulé un comportement « normal » pour les utilisateurs. Un utilisateur moyen a un profil temporel calqué sur une journée : il se connecte plus souvent en journée qu'au milieu de la nuit. En plus de cela, il est raisonnable de penser qu'il existe une corrélation entre « user-agent » et adresse IP pour les utilisateurs « normaux ». Illustrons : Alice, une utilisatrice typique de notre réseau social se connecte deux fois par jour. À sa pause déjeuner, elle se connecte depuis son travail, et en soirée depuis chez elle. Elle apparaîtra deux fois dans nos données, avec deux adresses IP différentes, mais aussi deux « user-agent » différents (Alice ne va pas au travail avec son ordinateur personnel). Des utilisateurs qui apparaîtraient avec un unique « user-agent » provenant de multiples adresses IP (ou l'inverse) seront a priori suspects.

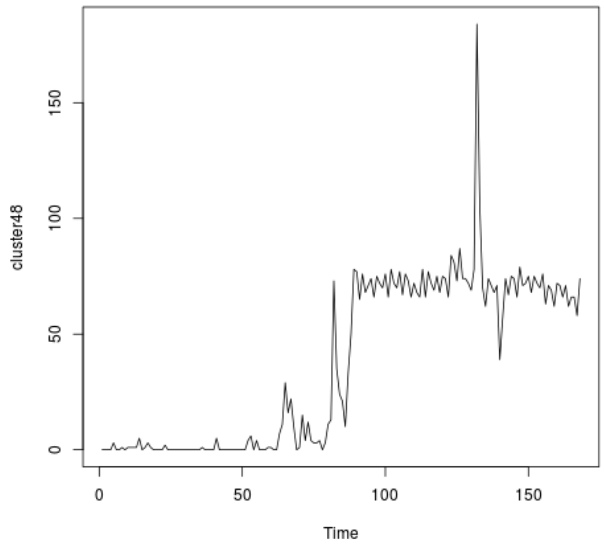
Pour valider ces deux hypothèses, j'ai tracé deux graphes (ci dessous) : le nombre de connexions par heure sur le réseau social, et la corrélation entre « user-agent » et adresse IP. Ces deux graphes utilisent toutes les données disponibles. Ils illustrent le comportement moyen des utilisateurs, que nous utiliserons comme témoin.





Le graphe temporel est peu surprenant : on peut lire les pics de connexion en journée, les périodes creuses la nuit. On peut même voir les week-ends (les deux derniers pics) au cours desquels les utilisateurs se connectent plus tard qu'en semaine. Le graphe de corrélation valide notre seconde hypothèse : on voit bien un pic apparaître pour une corrélation de 1 entre « user-agent » et adresse IP.

Fort de ces découvertes, j'ai ensuite tracé les graphes temporels et de corrélation pour les communautés trouvées précédemment. En voici un exemple :



Ces communautés d'utilisateurs ont été formées à l'aide d'adresses IP, sans utiliser de données temporelles. Le graphe temporel de cet agrégat dénote dramatiquement du comportement « normal » des utilisateurs : c'est un indice supplémentaire qui nous fait penser que ces comptes sont compromis.

Je n'ai malheureusement pas pu pousser mon travail plus loin. En effet, l'entreprise qui nous avait fourni les données n'a jamais répondu à nos relances. Nous aurions aimé avoir une confirmation du travail déjà fait ainsi que des données supplémentaires. À terme, l'objectif aurait été de tester nos méthodes de détection en direct, lors de la connexion de l'utilisateur et non plus après coup, en épluchant des logs.

Description factuelle : Sketchfab

J'ai commencé mon stage à Sketchfab à la mi février, une dizaine de jours après être rentré des États-Unis. L'équipe de Sketchfab était alors composé de trois personnes. Cédric, co-fondateur de l'entreprise, qui avait créé la première version du site et qui s'occupait de la partie « front-end » : Javascript, OpenGL etc. Alban, également co-fondateur, le « business developer » de l'entreprise. Pierre-Antoine, un ancien élève de Supélec qui avait rejoint l'équipe quelques mois plus tôt en tant que « CTO » (on fait tout à l'américaine dans le monde des start-oop), qui s'occupait plutôt de la partie « back-end » : Python, Django etc. C'est ce dernier qui allait m'encadrer dans mon stage. Je suis arrivé en même temps qu'un autre stagiaire, Clément, qui lui s'occupait plus de la partie 3D du site.

Ma première tâche a été de permettre aux utilisateurs de changer le fond d'écran de leurs modèles. Ils devaient pouvoir choisir parmi une liste d'images par défaut ainsi que de téléverser (moi je suis plutôt pour la francophonie m'voyez) les leurs. Ce travail a été très intéressant, quoiqu'un peu rude pour commencer, car il m'a fait découvrir toutes les parties du site, « front-end » comme « back-end ». Mes encadrants pensaient que j'aurai fini en un mois, il m'aura finalement fallu un peu moins de deux mois pour réaliser ce premier objectif. Ce délai a été causé par ma prise en main des outils et des pratiques de Sketchfab. J'ai également rencontré de nombreuses difficultés non prévues.

En tout cas, cette première mission a été très intéressante car horizontale. J'ai pu découvrir toutes les parties de l'application (sauf la partie 3D, pour laquelle j'étais peu qualifié et qui m'intéressait moins) et m'habituer à travailler avec l'équipe. J'ai aussi pu me rendre compte que je préférais travailler sur la partie Python Django que sur la partie Javascript, impression que j'ai partagée avec Pierre-Antoine. C'est pour cela que mes missions ultérieures ont comporté peu de Javascript.

Les trois « founders » de Sketchfab sont partis à New York à partir de la fin du mois de mars, à peine deux mois après mon arrivée, pour participer à l'incubateur Techstars. Je suis donc resté avec Clément dans les bureaux parisiens. La distance géographique n'a pas été un trop gros problème puisque nous nous parlions très régulièrement sur IRC ou même Skype. Les six heures de décalage entre New York et Paris se sont par contre fait ressentir : impossible d'appeler à l'aide avant 14h et soirées qui durent, qui durent, qui durent.

J'ai ensuite travaillé sur diverses « features », plutôt du côté de Django. De nombreuses optimisations, notamment dans les relations avec les bases de données, des envois de mail automatiques, un système d'invitation etc. Ces nombreuses tâches m'ont permis de m'améliorer énormément dans mon rapport avec Django, jusqu'à en avoir une importante maîtrise.

Un de mes plus gros projets aura été un réusinage massif du code Django accompagné du passage à Django 1.5. Le code Django originel avait été écrit par Cédric avec un objectif : que le site fonctionne rapidement. De ce « sprint » nous avions hérité une importante dette technique. J'ai donc passé beaucoup de temps à réusiner le code, l'adapter à nos nouveaux standards, ajouter des tests unitaires et corriger les rares erreurs existantes. En plus de cela, le site devait migrer sur la version suivante de Django. Cette nouvelle version cassait en partie la rétro-compatibilité : nous devions revoir le modèle d'utilisateur, central tant dans le code qu'en base de données. Je me suis donc frotté aux scripts Bash, aux transactions SQL et aux migrations critiques.

Lors de la fin de mon stage, j'ai surtout réalisé des « proof of concept ». Utilisant divers nouveaux outils à tester, pour la nouvelle version de l'API du site par exemple, mais aussi pour des systèmes de files de tâches, ou « task queue ». Comme j'étais le seul à avoir travaillé sur certaines parties critiques de l'application, j'ai également eu à former mes successeurs, eux aussi stagiaires (cherchez l'erreur), et à écrire de la documentation sur les bonnes pratiques que j'avais développées.

Ressenti, analyse : Seclab

L'arrivée aux États-Unis, et encore plus au Seclab, m'a un peu donné le vertige. On est seul, dans un pays où on ne parle pas français, où tout le monde conduit des grosses voitures et porte des armes, ça impressionne forcément un peu. Et puis on arrive dans une université américaine gigantesque, on s'y perd, on y prend son premier coup de soleil. Enfin on arrive au Seclab, on est entourés par des chercheurs plus qualifiés que nous, plus âgés, plus expérimentés bref, le vertige.

J'ai tout de même été gâté par toutes les personnes à qui j'ai parlé ou auxquelles j'ai demandé de l'aide car elles ont toutes été charmantes. Il n'empêche que je n'ai jamais osé déranger mes collègues pour les questions triviales que je me posais (« euh comment on ssh sur un serveur ? »), j'ai préféré chercher par moi-même sur Internet. Cette plongée dans le grand bain m'aura été très formatrice, j'ai appris à me débrouiller seul très rapidement.

J'ai été impressionné par les moyens disponibles dans le laboratoire. Le premier jour on m'a donné un bureau, un ordinateur (très puissant) et deux écrans vingt-deux pouces. Tous les chercheurs et les stagiaires étaient équipés de la sorte. Quand j'ai eu à construire des graphes énormes, après avoir expliqué que les huit giga de RAM de mon ordinateur ne suffisaient pas, on m'a tout de suite donné accès à un serveur avec quatre-vingt seize (96) giga de RAM. En plus de cela, j'aurais été payé près de mille huit cent dollars par mois, en plus des congés payés. L'université a clairement les moyens de ses ambitions.

Au niveau de l'organisation du travail, j'ai été complètement libre. La seule contrainte, une fois par semaine lors d'une réunion d'une petite heure, expliquer où j'en étais, les problèmes que j'avais rencontrés, etc. À part ça, libre à moi de rester chez moi toute la semaine pour y travailler (ou pas). J'ai beaucoup apprécié cette liberté car je supporte assez peu que l'on soit derrière mon dos à surveiller tous mes faits et gestes. Au demeurant, j'aurais tout de même vraiment gagné à prendre une petite remontrance de temps à autres, quand je lâchais la bride.

L'expérience m'aura été très favorable car j'ai pu découvrir le monde de la recherche, aux États-Unis du moins. J'ai pu me rendre compte que j'ai un vrai besoin de travailler au sein d'une équipe pour pouvoir donner mon maximum. L'interaction avec les autres m'aide énormément à progresser, à trouver des solutions à mes problèmes. Au Seclab, je n'avais pas vraiment d'interlocuteur direct que je pouvais embêter avec mes questions, ou de collègue avec qui réfléchir et cela m'a manqué.

Je suis aussi déçu de ne pas avoir pu terminer mon projet, à cause du manque de répondant du partenaire industriel. J'aurai perdu près de deux mois à attendre des données supplémentaires, ce sont deux mois que j'aurais pu mettre à profit.

Ressenti, analyse : Sketchfab

Contrairement à mon stage précédent, je suis arrivé plus confiant à Sketchfab. D'abord parce que j'étais de retour en France et que je n'avais plus le sentiment d'être seul, mais également parce que j'avais confiance en les compétences que je venais de développer. Cela ne m'a pas empêché d'apprendre énormément, bien sûr.

Après le luxe de l'université de Santa Barbara, arriver dans un open-space occupé par des start-ups parisiennes a été un petit choc. Fini l'ordinateur fourni par l'employeur, bienvenue l'ordinateur personnel recyclé en ordinateur professionnel (en mode « BYOM » ou « Bring Your Own Malware »). Fini également le calme. Tout le monde autour discute, négocie, mange etc. Plus de climatisation non plus, qui m'aura beaucoup manqué lors des chaleurs du mois de juillet. Et surtout, heureusement que je n'avais pas de loyer à payer (merci beau-papa), car ce n'est pas avec la gratification que je recevais que j'aurais pu vivre à Paris.

L'organisation de l'équipe était également très différente. J'avais beaucoup moins de liberté, on attendait de moi des résultats à (très) court terme, le travail était plus critique. J'étais plus cadré mais aussi mieux entouré. Travailler au sein d'une équipe, avec des interlocuteurs disponibles m'a vraiment servi, d'abord lors de ma période de formation mais également plus tard, quand j'ai eu à résoudre des problèmes ardus. Je me suis rendu compte que le simple fait d'avoir à expliciter clairement un problème apportait parfois une réponse. Il m'est souvent arrivé de poser une question à un collègue et de trouver la réponse en m'entendant parler. C'est quelque chose que je m'efforce désormais à faire, même quand je travaille seul.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec a diam lectus. Sed sit amet ipsum mauris. Maecenas congue ligula ac quam viverra nec consectetur ante hendrerit. Donec et mollis dolor. Praesent et diam eget libero egestas mattis sit amet vitae augue. Nam tincidunt congue enim, ut porta lorem lacinia consectetur. Donec ut libero sed arcu vehicula ultricies a non tortor. Ce paragraphe va-t-il passer inaperçu ? C'est une très bonne question. En attendant cela me permet de combler cette page qui ferait un petit peu peau de chagrin sans. Grâce à ce paragraphe, le saut de page qui vient n'est pas aussi désagréable à l'oeil qu'avant. N'hésitez pas à me dire ce que vous en pensez.

Le monde des start-ups, que je ne connaissais pas du tout a été une découverte très intéressante.

D'une part parce que, travaillant dans une petite équipe avec peu de personnes, nous avons tous eu à apprendre de nouveaux outils, de nouvelles méthodes. Nous avions également peu de temps pour réaliser nos objectifs. Ce sentiment d'adversité, quand il est partagé par toute l'équipe, est vraiment entraînant, presque grisant.

D'autre part, cet environnement est très différent ce celui des grandes entreprises traditionnellement présentées en école d'ingénieur. Tant dans l'organisation de l'équipe et des interactions avec ses collègues que dans les méthodes de développement utilisées. On nous a présenté lors du cours de Génie Logiciel les méthodes « classiques » de gestion de projet : les cycles en « V », les spécifications fonctionnelles de 200 pages et j'en passe et des meilleures. Ici tout est « agile » et, même si le terme est utilisé à outrance, il dénote de pratiques flexibles très intéressantes, que j'ai pu approfondire avec des livres comme « Rework » de Jason Fried et David Heinemeier Hansson. Alors que ma première rencontre avec le développement logiciel à Supélec m'avait laissé de marbre, cette re-découverte via un autre point de vue m'a passioné.

En plus de m'ouvrir à ces méthodes de développement, je me suis également intéressé à des langages de programmation dont on parle criminellement peu à Supélec, les langages fonctionnels. Nommément, j'ai lu le SICP, qui était le cours d'introduction à l'informatique au MIT. Livre très intéressant, encore aujourd'hui, parce qu'ils présentent l'informatique comme autre chose que seulement la programmation. Les exemples et exercices sont en SCHEME, un langage fonctionnel proche du LISP. J'ai également lu Real World Haskell, qui présente le langage Haskell (!). Même si je n'ai pas été amené à utiliser ces langages dans un environnement professionnel, leur connaissance m'apporte une ouverture importante que j'utilise dans n'importe quel langage.

Pour finir, j'ai aussi pu me rendre compte que les stagiaires sont vraiment très intéressants pour les entreprises, avant d'être des expériences formatrices pour les étudiants. Peut-être même plus dans une start-up, où il est « normal » d'attendre plus des employés. Ce ressenti m'a inspiré un petit billet d'humeur, qui me servira ici de conclusion.

Conclusion

J'ai été très enthousiasmé par ma césure et, comme je l'ai dit lors de la rentrée en troisième année, c'est pour l'instant ma meilleure année passée à Supélec. Ces stages m'ont permis de découvrir un monde et des compétences simplement inaccessibles depuis l'école. La découverte de l'entreprise m'a aussi conforté dans mes choix pour ma future carrière.

Je vais profiter de cette tribune pour remercier encore une fois Ludovic Mé, Christopher Kruegel et Giovanni Vigna, qui m'ont permis de passer six mois fantastiques à Santa Barbara.