

Rocquencourt, le 8 mars 2010

À qui de droit,

Yousra Chabchoub a effectué sa thèse sous ma co-direction (avec S. Fdida) entre 2006 et 2009 à l'INRIA Paris-Rocquencourt. Pendant ce travail, Yousra a exploré plusieurs questions importantes liées à la métrologie du réseau Internet. Ce domaine recouvre de nombreuses disciplines : les mesures proprement dites (installation de sondes, logiciels, ...), modélisation sous l'angle de la théorie du signal, de la théorie des probabilités, statistiques, conception d'algorithmes, validations, ... La très grande difficulté de cette discipline vient du fait qu'il n'y a pas modèle reconnu par une large communauté pour représenter le trafic TCP, comme c'est le cas du réseau téléphonique où les processus de Poisson sont admis par exemple.

Les travaux de thèse de Yousra ont été focalisés sur deux objectifs simples.

1. Donner une représentation robuste du trafic Internet.

Yousra a proposé une approche nouvelle qui consiste à faire des statistiques des flots sur une fenêtre de temps limitée. Si la durée  $W$  de cette fenêtre de temps est convenablement choisie, on obtient que la durée des flots suit une loi de Pareto dont le paramètre est remarquablement stable. L'originalité de l'approche vient du fait qu'un *algorithme*, i.e. sans pre-processing caché, détermine directement la quantité  $W$ . Cette approche a été validée sur un grand nombre de traces du trafic ADSL (Trafic Orange) et du trafic académique (Trafic Abilene aux États-Unis). Il y a, avec cette méthode, une perte d'information, on n'a plus la distribution de la longueur totale d'un flot mais seulement celle du nombre de paquets sur un intervalle de temps fixé. Le bénéfice de cette méthode qui contrebalance largement cette imprécision réside dans la robustesse des estimations obtenues, ce qui est le point faible (et même très faible) de la plupart des méthodes existantes (par un pre-processing plus ou moins avoué). Plus important, elle a montré, qu'en utilisant des résultats mathématiques qu'elle a développés, cette représentation partielle du trafic permettait d'obtenir des estimations robustes à partir d'un trafic échantillonné. L'échantillonnage est probablement le domaine où la robustesse fait le plus défaut, c'est donc un résultat de première importance.

2. Concevoir des algorithmes effectifs pour contrôler le trafic.

Si la description du trafic est un sujet important pour quantifier ses caractéristiques, la conception de réseau repose essentiellement sur un ensemble d'algorithmes qui permettent d'offrir un ensemble de fonctionnalités. Le thème général ici est de proposer des algorithmes adaptatifs, i.e. dont les paramètres très précisément identifiés sont calculés *automatiquement* par la procédure. (Pour une nouvelle fois bannir un pre-processing caché). Yousra a développé deux classes d'algorithmes, d'une part pour estimer le trafic échantillonné et d'autre part pour détecter de façon efficace et rapide les attaques de type "denial of service". Pour concevoir un algorithme dans ce domaine, il est important de bien comprendre quelles sont les constantes de l'algorithme qui dépendent du trafic (et donc de proposer une procédure pour les déterminer) et celles qui en sont indépendantes. C'est une source classique d'erreur et d'imprécision dans la littérature. À l'épreuve des traces de trafic, ses algorithmes ont montré leur très grande flexibilité, en ce qui concerne la détection d'attaques une phase de tests *en ligne* sur du trafic "réel" a été effectuée avec succès.

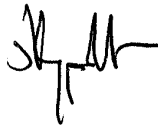
Ces travaux ont été effectués dans le cadre d'une collaboration avec Orange Labs et de projets RNRT. Les résultats obtenus par Yousra dans ces deux domaines ont été absolument remarquables. Elle a su combiner à la fois la modélisation et l'analyse avec la nécessaire (mais combien difficile) validation sur des ensembles

*significatifs* de données. Si ces travaux ont été fait dans le cadre d'une équipe, je tiens à souligner que plusieurs idées lui sont entièrement dues, notamment l'idée-clé d'introduire des statistiques locales pour estimer le trafic et de proposer l'algorithmique associée, ou encore pour les algorithmes de détection d'attaque, elle a été la principale contributrice. Yousra est une chercheuse très autonome. Elle a publié ses travaux dans des revues et des conférences internationales de premier rang.

Yousra a de plus de remarquables qualités de pédagogue, ses présentations orales sont en tout point excellentes : concision, précision et approfondissement variable en fonction de l'audience.

Pour conclure, je recommande très vivement, et avec enthousiasme, la candidature de Yousra. Je peux fournir plus de détails le cas échéant.

Bien cordialement,

A handwritten signature in black ink, appearing to read 'Philippe Robert', with a stylized, cursive script.

Philippe Robert

Professeur Georges HEBRAIL

### **Lettre de recommandation de Yousra Chabchoub**

Yousra Chabchoub est en Post-doc depuis Novembre 2009 à Télécom ParisTech au sein de l'équipe BILab, laboratoire de recherche commun entre Télécom ParisTech et EDF R&D sur le thème de l'informatique décisionnelle. Yousra intervient dans le cadre d'une collaboration avec EDF R&D. Son travail consiste à proposer de nouveaux algorithmes pour la gestion de flux de données. Un exemple d'application est de construire en temps réel des courbes agrégées (dans le temps et dans l'espace) à partir de nombreuses courbes individuelles de consommation électrique issues de sources distribuées. De telles informations sont très utiles pour la supervision et la maintenance des réseaux électriques. Yousra participe aussi à l'étude de l'architecture du réseau de collecte des mesures. L'objectif est de proposer une architecture robuste et fiable avec un calcul distribué qui supporte le passage à l'échelle.

Yousra fait preuve d'une bonne maîtrise de son domaine de recherche et de beaucoup d'autonomie. C'est une chercheuse sérieuse, ambitieuse et réactive. Ses présentations sont claires, pédagogiques et bien structurées.

Je recommande vivement la candidature de Yousra Chabchoub à un poste de Maître de Conférences des Universités.

Fait à Paris le 19 Mars 2010,





Christine Fricker  
INRIA, Domaine de Voluceau,  
Rocquencourt B.P. 105  
78153 le Chesnay Cedex

Rocquencourt, le 16 mars 2009

A qui de droit

Madame, Monsieur,

Yousra Chabchoub a effectué sa thèse, dans l'équipe-projet RAP (Réseaux, Algorithmes, Probabilités) de l'INRIA Paris-Rocquencourt dans laquelle je travaille, entre 2006 et 2009 sous la direction de Philippe Robert.

J'ai participé activement à son encadrement sur des problèmes liés à la métrologie d'Internet, sur laquelle une partie de l'équipe était investie. Ses travaux de thèse ont porté sur deux sujets:

- la détermination des caractéristiques du trafic Internet, et ceci même si la connaissance du trafic est partielle, dans le cas où le trafic est échantillonné.
- un algorithme d'identification des grands flots par filtrage (filtres de Bloom): son expérimentation à grande échelle (c'est-à-dire sur du trafic commercial fourni par Orange), l'analyse de l'algorithme, et l'adaptation de celui-ci pour la détection d'attaques (par déni de service ou en volume, les attaques les plus répandues sur le réseau).

Dans les deux cas, il s'agissait de développer des algorithmes en ligne adaptatifs. Plus précisément, ils devaient pouvoir s'adapter aux différences, mais aussi aux variations brusques ou lentes de trafic, et ceci sans qu'aucun pré-requis sur le trafic ne soit nécessaire.

Sur le premier sujet, Yousra a elle-même conçu un algorithme pour avoir des statistiques locales sur les flots; pour le deuxième, elle a conçu l'algorithme de détection d'attaques et a participé à l'analyse mathématique de l'algorithme.

Pour l'analyse mathématique, le problème a été simplifié. Il s'agissait d'estimer la probabilité de faux positifs, c'est-à-dire la proportion de flots courts prises pour des flots longs, dans cet algorithme probabiliste où l'on a une proportion d'erreurs. Vu le nombre élevé de flots courts, contrôler cette quantité permet d'estimer la performance de l'algorithme. Cette question était une question difficile conduisant à des résultats asymptotiques quand le nombre de cases du filtre est grand.

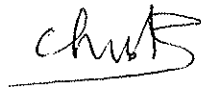
Yousra a participé pleinement à ce travail. Outre les discussions, nous lui devons ses simulations, ses interprétations, et ses tests pour une version modifiée de l'algorithme qui s'est avérée performante expérimentalement et incontournable pour conduire à un modèle qui puisse être analysé. Nous en avons rédigé deux articles avec des collaborations différentes, soumis à deux conférences internationales.

Yousra a fait preuve de beaucoup de qualités lors de ces collaborations: elle est capable d'autonomie, possède un bon sens physique qui lui permet d'intervenir pertinemment avec des collaborateurs même sur des aspects très mathématiques. Elle sait explorer avec succès des voies prometteuses et discuter avec des experts très variés. En effet, nos travaux ont regroupé des industriels (Orange Labs), des probabilistes et des algorithmiciens.

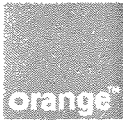
Yousra est réactive, avec toujours une bonne intuition et des remarques ou des questions pertinentes, qui font avancer la recherche. Elle sait établir des collaborations, a un excellent contact, ce qui en fait une personne agréable. Travailler avec elle était en conclusion un plaisir.

Elle sait aussi exposer en s'adaptant au public ses travaux dans un style simple et très clair. En ce sens, je crois en ses qualités de pédagogue, doublées de qualités humaines, qui la feront appréciée de ses étudiants.

Bien cordialement

A handwritten signature in black ink, appearing to read 'chris', with a horizontal line underneath.

Christine Fricker



**Fabrice GUILLEMIN**  
**Orange Labs**  
**Recherche et Développement**  
2, avenue Pierre-Marzin  
22307 Lannion Cedex  
0296051346  
fabrice.guillemine@orange-ftgroup.com

Référence : FT/RD/CORE/10/03/156

Lannion, le 18 mars 2010

A qui de droit.

Yousra Chabchoub a effectué sa thèse sous la direction de Philippe Robert, responsable du projet RAP à l'INRIA. Etant un collaborateur régulier du projet, c'est à cette occasion que j'ai pu rencontrer Yousra. Sa thèse portait sur la métrologie du trafic Internet et plus particulièrement sur l'inférence des caractéristiques des flots de trafic après échantillonnage déterministe et sur la détection d'anomalies de trafic en utilisant des filtres de Bloom. Ce dernier point a constitué la contribution du projet RAP au projet ANR OSCAR dont j'assurais la coordination technique.

Tout au long de sa thèse, Yousra a su concilier travaux théoriques et pratiques. En dehors de la conception d'algorithmes d'inférence statistique et de détection d'anomalies, Yousra a mené un travail expérimental très important sur des grandes masses de données. La validation pratique des algorithmes requiert en effet la manipulation de mesures de trafic issues de réseaux réels, en l'occurrence des réseaux d'Orange (réseaux de collecte ADSL et de transit international). Ces mesures se présentent sous forme de traces de trafic anonymes représentant quelques heures de trafic mais pouvant atteindre de très grands volumes (plusieurs centaines de Gigaoctets). La manipulation de telles masses avec des algorithmes exige un savoir faire particulier (hachage, filtrage, etc.), que Yousra maîtrise parfaitement. Le point fort de sa thèse était justement de montrer que les algorithmes étudiés résistaient à l'analyse en temps réel de grandes masses de données.

Yousra a montré au cours de sa thèse qu'elle maîtrise parfaitement de nombreux outils pratiques et théoriques, ce qui est en fait un élément de valeur dans une équipe de recherche. Je l'encourage donc à poursuivre sa carrière de recherche en milieu académique ou industriel.