

Numéro de poste :4073
Intitulé du poste :
Assistant-professorship position in Statistics and Data Mining
FICHE-RÉSUMÉ
CAMPAGNE DE RECRUTEMENT MCF 2012
UNIVERSITÉ PARIS DIDEROT

SECTION CNU de la candidate ou du candidat n° 27 (facultatif)

ETAT CIVIL

NOM : Chabchoub
Nom de jeune fille : Haddar
Prénom : Yousra
Age : 30 ans
Adresse professionnelle : ISEP, 21 rue d'Assas 75006 paris
Adresse email : yousra.chabchoub@gmail.com

ETUDES UNIVERSITAIRES

Année et Université où a été obtenu votre Master (DEA) :) 2006 – UPMC (Paris 6)
Année de soutenance de la Thèse et mention éventuelle : 2009, Très honorable
Titre de la thèse : Analyse et modélisation du trafic Internet
Ecole Doctorale et Université : EDITE – UPMC (Paris 6)
Discipline : informatique
Thèse financée ☒ OUI ☐ NON
Si oui quel financement ? Bourse INRIA

STAGE POST-DOCTORAL

1) Laboratoire (nom, adresse, Université) : BILab (Laboratoire d'informatique décisionnelle), Télécom Paris-Tech Département Infres, 46 Rue Barrault 75013 Paris
Période: de Novembre 2009 à Septembre 2010
Thème : Gestion de flux de données, en collaboration avec EDF R&D
Ayant conduit à : 2 publications dans 2 conférences internationales avec comité de lecture

PARTICIPATION À L'ENSEIGNEMENT

Monitorat (Université, période, discipline)

Nature et volume horaire des enseignements :

ATER (Université, période, discipline)

Vacations et enseignements à l'ISEP:

Lieu	Année	Discipline	Volume horaire
ISEP	2010-212	Cours et TP java	120
ISEP	2010-2012	Apprentissage par projet informatique	50
ISEP	2010-2012	TP Réseaux	36
ISEP	2010-2012	Algorithmes avancés	6
Télécom ParisTech	2008-2009	TP Réseaux	15
EFREI	2007-2008	TP langage C	42
UVSQ	2008-2010	TD mathématique	72
IUT Vélizy	2008-2009	TP Réseaux	60

Nature et volume horaire des enseignements :

Pour les **mutations** : maître de Conférence (Université, période, discipline)

RESUME DU PROJET DE RECHERCHE

Le domaine de gestion de flux de données suscite un intérêt particulier dans les années à venir vue la volumétrie sans cesse plus importante des données produites par les systèmes de mesure et les systèmes informatiques. On est aujourd'hui face à une multitude d'objets communicants qui envahissent notre quotidien. Ces objets sont capables de faire des mesures et de s'échanger des données.

Afin de réduire le volume de données à traiter, deux axes méritent d'être approfondis:

1/ L'échantillonnage: On a en particulier besoin d'un échantillonnage sélectif qui permet de construire, à moindre coût, un échantillon significatif, adapté à l'objectif final de l'application. Dans la littérature, on trouve très peu de travaux de recherche qui proposent un tel échantillonnage.

2/ Les algorithmes basés sur des sketches: Un sketch est un résumé compact des données basé sur des projections aléatoires via des fonctions de hachage. Les sketches sont très utilisés dans le domaine des flux de données car ils assurent un traitement rapide en utilisant une mémoire de petite taille. La détection des anomalies par exemple est partiellement basée sur les sketches.

La taille des systèmes et des réseaux impose aujourd'hui le recours à une architecture décentralisée pour plus de robustesse et de fiabilité. Certaines applications initialement conçues pour fonctionner dans un environnement complètement centralisé nécessitent une migration vers l'architecture décentralisée. Ceci demande une réflexion sur les informations à échanger, la répartition des tâches, la prise de décision...

Pour détecter les attaques dans le trafic Internet par exemple, on peut concevoir une architecture décentralisée basée sur des sondes actives réparties dans le réseau (au niveau de certains routeurs). La complémentarité entre ces points d'observation permettra d'avoir une vue globale du réseau. Ainsi, on pourra même détecter les

attaques distribuées qu'une seule sonde est incapable de relever par analyse de son trafic local uniquement.

PUBLICATIONS

	Nombre total
Articles acceptés dans des revues internationales à comité de lecture	2
Autres articles	
Chapitre(s) d'ouvrage / ouvrage	
Article(s) de vulgarisation	
Articles acceptés dans des conférences à comité de lecture	5

Liste des 3 publications majeures (titre, auteurs, journal)

1) A stable statistical estimation of Internet traffic with application to sampling. Yousra Chabchoub, Christine Fricker, Fabrice Guillemin, and Philippe Robert. Computer Communications Journal, 33(1) :103-112, 2009.

2) Analysis of an algorithm catching elephants on the Internet. Yousra Chabchoub, Christine Fricker, Frederic Meunier, and Danielle Tibi. In proceedings of the Discrete Mathematics and Theoretical Computer Science, Septembre 2008.

3) Yousra Chabchoub, Christine Fricker, Fabrice Guillemin, and Philippe Robert. Inference of flow statistics via packet sampling in the Internet. IEEE Communications Letters, 12(12) : 897 - 899, 2008.

PARTIE RESERVEE AUX RAPPORTEURS
Les candidats ne doivent rien écrire dans cette section

Nom du candidat :

Nom du rapporteur :

Appréciation de :

- La qualité du dossier (A+, A, B, C) :
- L'adéquation avec l'emploi (parfaite, correcte, faible) :

Rapport circonstancié :