Proposition de sujet de thèse CNRS-L/UPPA

2017-2018

CANA-CNRS pour la recherche marine au Liban

Dans le cadre de l'accord entre le Conseil National de la Recherche Scientifique de la République Libanaise (CNRS-L) et l'Université de Pau et des Pays de l'Adour (UPPA) pour le co-financement des thèses de doctorat dans des thématiques d'intérêt commun, **trois bourses de recherches doctorales pour l'année 2017-2018** seront mises en place. Ces thèses sont proposées conjointement par un laboratoire de recherche de l'UPPA et un laboratoire de recherche libanais dans le cadre d'une convention de co-tutelle ou de co-direction. Ainsi, les équipes souhaitant proposer des thèses de doctorat pour l'année 2017-2018 sont priées de compléter ce formulaire de proposition de sujet de thèse et de l'envoyer par courriel **avant le 11 septembre 2017** à: tamara.elzein@cnrs.edu.lb (pour CNRS Liban) et jacqueline.petitbon@univ-pau.fr (pour le collège des Ecoles doctorales de l'UPPA). **Les sujets retenus seront diffusés pour l'appel à candidature, et la sélection finale des boursiers se fera par un comité mixte des deux institutions.**

Il est à noter que les thématiques prioritaires pour l'année 2017-2018 sont les suivantes :

- **Ressources aquatiques**
- **Géophysique/géo-ressources**
- **Archéologie/archéométrie**
- **Géographie/aménagement/ télédétection**
- **Eco-construction**
- **Durabilité des ouvrages**
- **Environnement**
- **Energie**
- **Matériaux**
- **Informatique**

Pièces à joindre :

- CV du co-directeur libanais
- CV du co-directeur français

II. Fiche de Renseignements sur le laboratoire d'accueil au Liban

Université ou centre de recherche **: Université Antonine**
Laboratoire d'accueil : **TICKET Lab**
Nom du Directeur du laboratoire : **Bechara Al Bouna, PhD**

Adresse : **Hadat- Baabda, Liban**

Ville : **Baabda**

Tél./Fax/Mél :

Tél.:+ 961 5 92 70 00
Fax:+ 961 5 92 70 01

Faculté ou organisme auquel est affilié le laboratoire d'accueil : **Université Antonine - Faculté d'Ingénieurs en Informatique, Multimédias, Réseaux et Télécommunications**.

Nom du Directeur de thèse : **Bechara Al Bouna, PhD**

Le Directeur de thèse fait-il partie du laboratoire d'accueil : **x** Oui / □ Non

Si non, précisez son rattachement et ses coordonnées :

- Principaux thèmes de recherche de l'équipe où sera effectué le travail de thèse :
  - Sécurité, control d'accès et anonymisation de données
  - Optimisation des données
  - Management des systèmes d'information

- Liste des publications récentes de l'équipe (pertinentes au sujet proposé- 3 dernières années) :

  - Mohamed Nassar, Nathalie Wehbe, Bechara al Bouna: K-NN Classification under Homomorphic Encryption: Application on a Labeled Eigen Faces Dataset. CSE/EUC/DCABES 2016: 546-552.
  - Sara Barakat, Bechara al Bouna, Mohamed Nassar, Christophe Guyeux: On the Evaluation of the Privacy Breach in Disassociated Set-valued Datasets. SECRYPT 2016: 318-326.
  - AL Bouna B, Raad EJ, Chbeir R, Elia C, Haraty R. 2015. Anonymizing multimedia documents. World Wide Web. :1–21.
  - AL Bouna B, Clifton C, Malluhi Q. 2015. Anonymizing transactional datasets. Journal of Computer Security. 23:89–106.
  - AL Bouna B, Clifton C, Malluhi QM. 2015. Efficient Sanitization of Unsafe Data Correlations. Proceedings of the Workshops of the {EDBT/ICDT} 2015 Joint Conference (EDBT/ICDT), Brussels, Belgium, March 27th, 2015. :278–285.
  - AL Bouna B, Couchot J-F, Couturier R, Fadil YAhmed, Guyeux C. 2015. Performance Study of Steganalysis Techniques.Applied Research in Computer Science and Engineering (ICAR), 2015.
  - Raad E, AL Bouna B, Chbeir R. 2015. Preventing sensitive relationships disclosure for better social media preservation.International Journal of Information Security. :1–22.

- Ramzi A. Haraty, Mohammad A. Taha, Bechara al Bouna: MOP: A Privacy Preserving Model for Multimedia Objects. MEDES 2014: 213-219

La thèse sera-t-elle effectuée en co-tutelle ou co-direction: **en co-direction**

III. Fiche de Renseignements sur le laboratoire d'accueil à l'UPPA

Laboratoire d'accueil : **LIUPPA**

Nom du Directeur du laboratoire : **Pr. Richard Chbeir**

Adresse : **Avenue de l'Université BP 576 64012 PAU cedex**

Code postale-Ville : **64012**

Tél./Fax/Mél :

De France : **+ 33 5 59 57 43 37**

De l'étranger : **+ 33 5 59 57 43 37**

Ecole doctorale auquel est affilié le laboratoire d'accueil : ÉCOLE DOCTORALE SCIENCES EXACTES ET LEURS APPLICATIONS - ED211

Nom du Directeur de thèse : **Pr. Richard Chbeir**

Le Directeur de thèse fait-il partie du laboratoire d'accueil : **x** Oui / □ Non

Si non, précisez son rattachement et ses coordonnées :

Nombre de thèses dirigées (ou co-dirigées) actuellement : 4 dont 1 en co-direction et 1 en co-tutelle

Pour les cinq dernières années, précisez les thèses soutenues, la durée en mois pour chacune d'entre elle, la liste des publications et la situation actuelle de chaque diplômé.

Etudiante :	Khouloud Salameh
Sujet :	Gestion de ressources dans un écosystèmes numérique : applications dans les smart grids
Durée :	40 mois
Taux de co-encadrement :	50%
Financement :	Contrat doctoral à l'UPPA (en cotutelle avec l'Université du Pays Basque -Espagne
Co-encadrants :	Haritza Camblong Ruiz à l'Université du Pays Basque -Espagne
Devenir de l'étudiante :	Post-Doc à l'UPPA
Publications :	

[1] Khouloud Salameh, **Richard Chbeir**, Haritza Camblong, Gilbert Tekli, Ionel Vechiu, *A Generic Ontology-Based Information Model for Better Management of Microgrids*. Proc. of the 11th International Conference Artificial Intelligence Applications and Innovations, AIAI 2015, Bayonne, France, September 14-17, 2015 : pp. 451-466 (taux d'acceptation de 43%)

[2] Khouloud Salameh, **Richard Chbeir**, Haritza Camblong, Ionel Vechiu, A Multi-objective Cooperative Model: An Application on Microgrids, IEEE Transactions on Sustainable Computing, 2017 (A paraitre)

[3] Khouloud Salameh, **Richard Chbeir**, Haritza Camblong, Multi-objective Cooperative Scheduling for Smart Grids, 25th International Conference on COOPERATIVE INFORMATION SYSTEMS (COOPIS 2017), Rohdes-Greece, 2017 (A paraitre)

[4] Khouloud Salameh, **Richard Chbeir**, Haritza Camblong, Microgrid Components Clustering in a Digital Ecosystem Cooperative Framework, 21st International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2017), Marseille-France, 2017 (A paraitre)

Etudiante : Regina Paola Ticona Herrera
Sujet : Normalisation RDF
Durée : 42 mois
Taux de co-encadrement : 50%
Financement : Bourse du gouvernement Péruvien (LASPAU)
Co-encadrants : Joe Tekli à LAU – Liban, Sébastien Laborie à l'UPPA
Devenir de l'étudiante : Maitre de conférences à San Pablo Catholic University, Arequipa-Peru
Publications :

[5] Regina Ticona Herrera, Joe Tekli, **Richard Chbeir**, Sébastien Laborie, Irvin Dongo, Renato Guzman, *Toward RDF Normalization*, Proc. of the 34th International Conference on Conceptual Modeling (ER 2015), Stockholm, Sweden, October 19-22, 2015: pp.261-275 (taux d'acceptation de 16%)[1]

Etudiante : Eliana Raad
Sujet : Gestion de la vie privée dans les réseaux sociaux
Durée : 38 mois
Taux de co-encadrement : 100%
Financement : Allocation Ministérielle
Devenir de l'étudiante : ATER à l'UPPA
Publications :

[6] Bechara al Bouna, Eliana J. Raad, **Richard Chbeir**, Charbel Elia, Ramzi A. Haraty, *Anonymizing multimedia documents*, World Wide Web Journal, 19(1): 135-155 (2016)

[7] Eliana J. Raad, **Richard Chbeir**, *Foto2Events: From Photos to Event Discovery and Linking in Online Social Networks*, Proc. of the IEEE 4th International Conference on Big Data and Cloud Computing, BDCloud 2014, Sydney, Australia, December 3-5, 2014: pp. 508-515

[8] Bechara al Bouna, Eliana J. Raad, Charbel Elia, **Richard Chbeir**, Ramzi A. Haraty, *de-linkability: a privacy-preserving constraint for safely outsourcing multimedia documents*. Proc. of the 5th ACM International Conference on Management of Emergent Digital EcoSystems (MEDES '13), Luxembourg, Luxembourg, October 29-31, 2013: pp. 68-75 (taux d'acceptation de 30%)

[9] Elie Raad, **Richard Chbeir**, Albert Dipanda, Eliana J. Raad: Automatic rule generation using crowdsourcing for better relationship type discovery. Pervasive and Mobile Computing 36: 80-97 (2017)

Etudiant : Solomon Asres Kidanu
Sujet : Digital Ecosystem: For Better Management of Multimedia Contents
Durée : 42 mois
Taux de co-encadrement : 100%
Financement : Bourse Ambassade de France en Ethiopie
Devenir de l'étudiant : Enseignant-chercheur à l'Université d'Addis-Abeba en Ethiopie

[1] Nous attendons la notification d'un papier soumis au journal Journal of Web Semantics (Resolving Logical Redundancies and Physical Disparities in RDF Descriptions)

Publications :

[10] Solomon Asres Kidanu, Yudith Cardinale, Gilbert Tekli, **Richard Chbeir**, *A Multimedia-Oriented Digital Ecosystem: A new collaborative environment*, Proc. of the 14th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2015, Las Vegas, NV, USA, June 28 - July 1, 2015: pp. 411-416

[11] Solomon Asres Kidanu, Yudith Cardinale, **Richard Chbeir**, Victor De Ponte, Alejandro Figueroa, Ronier Rodríguez, and Carlos A. Raymundo Ibanez, *MMDES: Multimedia Digital Ecosystem - New Platform for Collaboration and Sharing,* Proc. of the 19th IEEE International Conference on Computational Science and Engineering (CSE 2016) (A paraître)

[12] Corentin Donzelli, Solomon Asres Kidanu, **Richard Chbeir**, Yudith Cardinale: Onto2MAS: An Ontology-Based Framework for Automatic Multi-Agent System Generation. SITIS 2016: 381-388

Etudiante : Mônica Ribeiro Porto Ferreira
Sujet : Optimizing similarity queries in metric spaces meeting user's expectation
Durée : 36 mois
Taux de co-encadrement : 50%
Financement : Bourse COFECUB
Co-encadrants : Caetano Traina Junior (Thèse en cotutelle avec l'Université Sao Paolo - Brésil)
Devenir de l'étudiante : Analyste de Données à Serasa Experian – Brésil

Publications :

[13] Mônica Ribeiro Porto Ferreira, Lucio F. D. Santos, Agma J. M. Traina, Ires Dias, **Richard Chbeir**, Caetano Traina Jr., *Algebraic Properties to Optimize kNN Queries*, JIDM 2(3): 385-400 (2011)

[14] Mônica Ribeiro Porto Ferreira, Marcela Xavier Ribeiro, Agma J. M. Traina, **Richard Chbeir**, Caetano Traina Jr., *Adding Knowledge Extracted by Association Rules into Similarity Queries*, JIDM 1(3): 391-406 (2010)

[15] Mônica Ribeiro Porto Ferreira, Marcelo Ponciano-Silva, Agma Juci Machado Traina, Caetano Traina Jr., Sandra De Amo, Fabiola S. F. Pereira, **Richard Chbeir**, *Integrating User Preference to Similarity Queries over Medical Images Datasets*, Proc. of the 23rd IEEE International Symposium on Computer-Based Medical Systems (CBMS 2010), 1, pp. 486-491, Perth, Australia, IEEE Computer Society, October 2010

[16] Mônica Ribeiro Porto Ferreira, Agma Juci Machado Traina, Ires Dias, **Richard Chbeir**, Caetano Traina Jr., "*Identifying Algebraic Properties to Support Optimization of Unary Similarity Queries*", 3rd Alberto Mendelzon International Workshop on Foundations of Data Management (AMW 2009), 450, pp. 1-10, Arequipa, Peru, CEUR-WS, May 2009

Principaux thèmes de recherche de l'équipe où sera effectué le travail de thèse :

Les principaux thèmes de recherche de l'équipe s'articulent autour de l'extraction et de la protection de l'information appliqués dans trois domaines : **Réseaux de capteurs, web intelligent** et **écosystèmes numériques**.

Liste des publications récentes de l'équipe (pertinentes au sujet proposé) :

[17] Elie Raad, Richard Chbeir, Albert Dipanda, Eliana J. Raad: Automatic rule generation using crowdsourcing for better relationship type discovery. Pervasive and Mobile Computing 36: 80-97 (2017)

[18] Bechara al Bouna, Eliana J. Raad, Richard Chbeir, Charbel Elia, Ramzi A. Haraty, *Anonymizing multimedia documents*, World Wide Web Journal, 19(1): 135-155 (2016)

[19] Elie Raad, Bechara al Bouna, Richard Chbeir: Preventing sensitive relationships disclosure for better social media preservation. Int. J. Inf. Sec. 15(2): 173-194 (2016)

[20] Elie Raad, Richard Chbeir, Albert Dipanda: Discovering relationship types between users using profiles and shared photos in a social network. Multimedia Tools Appl. 64(1): 141-170 (2013)

[21] Elie Raad, Richard Chbeir: Privacy in Online Social Networks. Security and Privacy Preserving in Social Networks 2013: 3-45

[22] Eliana J. Raad, Richard Chbeir, *Foto2Events: From Photos to Event Discovery and Linking in Online Social Networks*, Proc. of the IEEE 4th International Conference on Big Data and Cloud Computing, BDCloud 2014, Sydney, Australia, December 3-5, 2014: pp. 508-515

[23] Bechara al Bouna, Eliana J. Raad, Charbel Elia, Richard Chbeir, Ramzi A. Haraty, *de-linkability: a privacy-preserving constraint for safely outsourcing multimedia documents*. Proc. of the 5th ACM International Conference on Management of Emergent

Digital EcoSystems (MEDES '13), Luxembourg, Luxembourg, October 29-31, 2013: pp. 68-75 (taux d'acceptation de 30%)
[24] Bechara al Bouna, Richard Chbeir, Alban Gabillon, Patrick Capolsini: A Flexible Image-Based Access Control Model for Social Networks. Security and Privacy Preserving in Social Networks 2013: 337-364
[25] Bechara al Bouna, Richard Chbeir, Alban Gabillon, Patrick Capolsini: A Fine-Grained Image Access Control Model. SITIS 2012: 603-612
[26] Elie Raad, Richard Chbeir, Albert Dipanda: Rules, Photos, and Crowdsourcing for Relationship Type Discovery in Social Networks. JMPT 2(2): 90-122 (2011)
[27] Bechara al Bouna, Richard Chbeir, Alban Gabillon: The Image Protector - A Flexible Security Rule Specification Toolkit . SECRYPT 2011: 345-350
[28] Elie Raad, Richard Chbeir, Albert Dipanda: User Profile Matching in Social Networks. NBiS 2010: 297-304
[29] Bechara al Bouna, Richard Chbeir, Stefania Marrara: Enforcing role based access control model with multimedia signatures. Journal of Systems Architecture - Embedded Systems Design 55(4): 264-274 (2009)

IV. Sujet de thèse

A faire signer obligatoirement par tous les co-directeurs

IV.1. Titre

Differentially Private Image Classification

*La thèse fait-elle partie d'un projet de recherche financé par le CNRS-L : □ Oui / **x** Non

Si oui, précisez :

*La thématique sous laquelle s'inscrit la thèse fait-elle partie des priorités de cet appel pour l'année 2017-2018 (voir annexe): **x** Oui / □ Non

Si oui, précisez (possibilité de choisir plus qu'une) : **Informatique**

Si non, définir une:

IV.2. Résumé (ne pas dépasser 200 mots)

In this thesis, our aim is to design and develop an anonymous full-duplex image classification service under Differential Privacy. We work under the assumption that both, the cloud and the querier are semi-trusted entities, thus their data should remain safe and confidential. That is, neither the querier nor the cloud should be able to link a particular individual to an image on the cloud while maintaining, to a certain extent, suitable classification accuracy. We use Principal Component Analysis (PCA) to transform sample images into anonymized vectors; differentially private synopsis of PCA vectors, and we ensure that these vectors remain unidentifiable.

IV.3. Contexte et problématique (ne pas dépasser 200 mots)

The research question of this project is how to get benefit from the statistical information of the collected data while not harming the individual privacy of any participant in the collection of the data?

Preserving privacy is not an easy task, and things do not end well in most of the scenarios. For example, in year 2000, Netflix released an "anonymized" movie viewing dataset by stripping all identifying information. They wanted to make the dataset available for enhancing their movie recommendation algorithms. Unfortunately, their de-identification was vulnerable. Narayanan and

Shmatikov [12] showed that they can re-identify specific users, and predict their political affiliation. All they needed is some small extra amount of information about a given user. Differential Privacy is a set of tools that was designed to provide a strong and robust anonymization scheme.

In our scenario, we want to benefit from the scalability, the ubiquity and the efficiency of the cloud when comes to big data applications. We want to run a service for image classification in a differentially private manner. This can help anonymous biometric authentication or localization.

“Differential privacy” describes a promise, made by a data holder, or curator, to a data subject: “You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, datasets, or information sources, are available.” [13]. Differential Privacy ensures that the probability that an aggregate query will produce a given result is almost the same if it’s conducted on two datasets that differ by only one record.

In this thesis, we assume that a party that holds a database of images. The images belong to different classes and each image is labeled with its class number. Another party has an image of unknown class and wants to predict this class. In non-private settings, this would be easy: We train a classifier based on the labeled image database and we predict the unknown image based on the learnt classifier. The second party shares the image with the first party, or may be the first party sends the learnt model to the second one. Both scenario may expose private data. For example, if we use a support vector machines model, the model itself is made of support vectors, which represent individual records.

In private settings, we want to imagine a man in the middle that asks information about the image, and asks information about the image database (or classification model).

Both parties must add noise (in a manner that we want to investigate in this project) to make the whole process differentially private. We want in result that any single image in the database is masked. The query image is masked as well.

One idea is to embed images into vector spaces using PCA and add noise to the PCA vectors. In the database side having the ground truth images, we want to test several filtering and sampling techniques and transform the database into a synopsis dataset of unidentifiable images. We want to test our approach using different classifiers such as K-nearest neighbors and kernel density estimation [5].

We also want to study the trade-off accuracy vs. privacy in terms of the so-called “privacy budget”. The privacy budget is the total allowed leakage as determined by the number of answered queries and the accuracy of the answers. The devil in the DP is in the privacy budget: If we set it too high, we leak your sensitive data. If we set it too low, then the answers we get might not be particularly useful.

IV.4. Descriptif des objectifs et de l’impact (ne pas dépasser 200 mots)

The project aims at providing safe image classification on the cloud. This is motivated by the fact that cloud computing is becoming nowadays essential for the success of many organizations. Cloud computing offers scalability, ubiquity, and efficiency, shaping the next generation of information technology, and taking to the next level, the development of end-user services such as location identification service; a sample scenario to develop in our project.

In this scenario, we assume that a querier communicates with a cloud-based location identification service to determine the places where some images were taken. Intuitively, the querier submits one or several images to the cloud service and retrieves back the recognized places from the images.

This is a typical classification scenario, but for privacy reasons, neither the querier, nor the cloud-based service should be able to recognize/identify individuals in the images. Both, the trained images stored at the cloud-based service, and the images provided by the querier must be anonymized to ensure privacy while at the same time, classification must remain accurate.

IV.5. Aspect appliqué et/ou aspect innovateur (ne pas dépasser 200 mots)

The idea is to ensure that the sample images stored on the cloud and the images in the request are both differentially private in such a way that the trade-off between accuracy and privacy remains acceptable.

Privacy preserving cloud-services rely primarily on encryption [4] to ensure that the data stored in their premises is safe and secure. Encryption however, presents large computational overhead and therefore, in this project, we opt for anonymization instead of encryption by adding noise to the PCA vectors to achieve differential privacy [6].

We use Kernel Density Estimation [5] and K-Nearest Neighbors [5] algorithms to perform private classification and compare the results of each of the algorithms on the synopsis sample of differentially private PCA vectors [6]. The classification occurs on the cloud to benefit from cloud computing processing. Our framework is composed of three main components:

- *Cloud service*: hosts the sample noisy images, which are used to train the classification algorithm.
- *Anonymization service*: converts images in the main dataset to a PCA vector and adds noise to the vectors to hide sensitive information. Performs sampling and filtering and returns a synopsis dataset.
- *Querying service* – Querier: submits requests to the cloud-service.

IV.6. Etat des recherches dans le domaine avant la thèse (ne pas dépasser 200 mots) + Ref. Bibliographiques

The K-Nearest Neighbors (K-NN) [9] is one of the most popular and influential data mining algorithms in the literature. However, many adaptive attacks can form a real threat to data privacy

in K-NN based systems. Li et al. propose in [10] a privacy-preserving system based on Kernel density estimation using Gaussian Kernel instead of K-NN.

Their system, as most of other prior work, uses computation over encrypted data. They describe four roles:

- The data owner submits encrypted data to the system.

- The querier submits encrypted queries to receive classification results.

- The host role, possessed by the cloud, stores the incoming encrypted data and hosts the classification.

- Finally, the Cryptographic Service Provider (CPS) owns both encryption and decryption key.

The main problem in this work is that the system relies on multiple mutually disturbed data owners where some of them are not completely trusted.

The work described in [7] turns K-means clustering algorithm to a differentially private algorithm, where noise are added to the centroids in a way that respects the requirements of Differential Privacy. Differentially private K-Means is divided into two approaches: interactive and non-interactive.

Interactive approach is based on a query that can be used just once, can serve only one querier and only for one task. Any mechanism that is based on this approach returns a noisy result to the user. Each query has a budget given by the database owner. Each execution makes the budget loses ε of its value. When the budget is less than ε, the query can't be executed anymore. Hence this approach has many restrictions when executing queries to provide any privacy breach.

The non-interactive approach algorithms return a noisy synopsis data set. The querier can send queries to this synopsis to get statistical noisy data. This approach has no limits or restrictions to the number and the sender of the queries.

The work in [1] [2] is similar to our proposed idea but relies on a partially homomorphic encryption called Paillier's encryption. This type of encryption is a public key scheme; it means that the encryption can be done by a public key while decryption can only be done by a trusted party that possesses the private key. The technique however assumes that the cloud is a trusted party, and thus the privacy of the dataset is threatened.

[1] C. Dwork. Differential Privacy. In ICALP, pages 1-12, 2006.

[2] C. Dwork. A firm foundation for private data analysis. Commun. ACM, 54(1), pages 86-95, Jan. 2011.

[3] K. Chaudhuri, A. D. Sarwate, K. Sinha. Near-optimal Differentially Private Principal Components, 2012.

[4] M. Nassar, N. Wehbe, B. Al Bouna. K-NN Face Classification under Homomorphic Encryption. In CSE, 2016.

[5] F. Liy, R. Shiny, V. Paxson. Exploring Privacy Preservation in Outsourced K-Nearest Neighbors with Multiple Data Owners. Proceedings of the 2015 ACM Workshop on Cloud Computing Security Workshop. ACM, 2015.

[6] Wuxuan Jiang, Cong Xie, Zhihua Zhang: Wishart Mechanism for Differentially Private Principal Components Analysis. AAAI 2016: 1730-1736

[7] D. Su, J. Cao, N. Li, E. Bertino, H. Jin. Differentially Private K-Means Clustering. CODASPY '16 Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy. ACM, 2016.

[8] F. D. McSherry. Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data (SIGMOD), pages 19–30, 2009.

[9] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg. Top 10 algorithms in data mining. Knowl. Inf. Syst., Dec. 2007.

[10] F. Liy, R. Shiny, V. Paxson. Exploring Privacy Preservation in Outsourced K-Nearest Neighbors with Multiple Data Owners. Proceedings of the 2015 ACM Workshop on Cloud Computing Security Workshop. ACM, 2015.

[11] Y. Elmehdwi, B. K. Samanthula, and W. Jiang. Secure k-nearest neighbor query over encrypted data in outsourced environments. In Proceedings of the IEEE International Conference on Data Engineering, ICDE'14.

[12] Narayanan, A., & Shmatikov, V. (2008, May). Robust de-anonymization of large sparse datasets. In 2008 IEEE Symposium on Security and Privacy (sp 2008) (pp. 111-125). IEEE.

[13] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4), 211-407.

[14]https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy/

IV.7. Programme de recherche prévu pour la thèse et contribution des différents partenaires (ne pas dépasser 200 mots)

We divided the work according to the following tasks:

1. Conduct a comprehensive literature review of the field and some research in the same domain. There are many techniques that need to be studied carefully before we develop the differentially private image scheme. Some of these techniques rely primarily on homomorphic encryption while others apply private classification on textual data.

2. Design the cloud-service architecture. We will also seek to anonymize principal component analysis vectors using differential privacy and prove that these vectors are de-identifiable. Both the principal and the co-investigators will work on these tasks to ascertain their viability, correctness and efficiency.
3. Implement our cloud-service at TICKET Lab. and produce a working prototype for our techniques. We plan on testing them, and demonstrate their effectiveness by producing experimental results.

IV.8. Calendrier prévisionnel des mobilités

Researcher	Destination	Planned date	Duration (Month)	Mission objectives
Student	France	Nov-17	6	**First visit:** Definition & discussion on databases to be analyzed Evaluation study of data Evaluation of project datasets
Richard Chbeir	Lebanon	Jun-18	1	Discussion et position paper writing
Student	France	Nov-18	6	Definition & discussion on experimental scenarios Comparison of proposed methods and results Evaluation and comparison of experimental results
Béchara Al Bouna	France	Jan-19	1	Discussion et paper writing
Student	France	Nov-19	6	Analysis of results and prospective improvements Assessment of prototypical system and implementations Supervision of prototypical system completion

IV.9. Diffusion/valorisation des résultats

The main outcome of this thesis will be the location identification cloud-service to be implemented at the servers of LIUPPA Lab and TICKET Lab. Since validation is a research requirement, evaluation results will be delivered through publications in the research literature.

We will elaborate several tests to evaluate:

1. The efficiency of our differentially private PCA vectors by ensuring that the anonymized data are consistent and non-identifiable.

2. The loss in the accuracy of the classifier for the sake of privacy.

3. The performance of the cloud-service compared to the approach proposed in [1] that uses encryption in its underlying scheme.

IV.10. Compétences requises

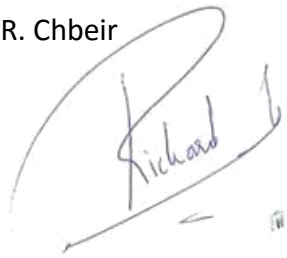
Basic knowledge in data anonymization and particularly differential privacy

Advanced knowledge in programming

Date : 11/09/2017

Noms et signatures (directeurs de thèse)

R. Chbeir



B. Al Bouna