

Table des matières.

Introduction générale.	I
Chapitre 1: Préservation de privacy en data mining.	
I. Introduction	2
II. Difficultés de définir la privacy.	3
III. Préservation de privacy en datamining.	4
III.1 Préservation de la privacy individuelle.	4
III.2 Préservation de la privacy collective.	4
IV. Carctérisation des scénarios dans la préservation de privacy en datamining.	5
V. Préservation de la privacy en datamining : Modèles et algorithmes.	6
V.1 La classification de V. S. Verykios et E. Bertino.	6
V.2 La classification de C. Aggarwal et S.YU.	7
V.2.1 La méthode de perturbation des données.	7
V.2.2 Les modèles <i>k-anonymité</i> et <i>l-diversité</i> .	7
V.2.3 Préservation distribuée de la privacy.	8
V.2.4 Rétrogradation de l'efficacité de l'application.	8
V.3 Autres classifications.	8
VI. Les métriques d'évaluation des algorithmes de préservation de privacy en datamining.	9
VII. Les modèles de distribution de données.	10
VIII. Conclusion.	13
Chapitre 2 : L'algorithme de clustering k-means.	
I. Introduction.	15
II. Définition formelle du clustering.	15
III. Les étapes de clustering.	15
III.1 Préparation des données.	15
III.2 Choix de l'algorithme de clustering.	16
III.3 Exploitation des clusters.	16
IV. Méthodes de clustering.	16
IV.1 Les structures produites (hiérarchies vs partitions).	16
IV.1.1 Clustering par partitionnement.	16
IV.1.2 Clustering hiérarchique.	17
IV.2 Méthodes agglomératives (ascendantes) vs divisives (descendantes).	17
IV.3 Clustering dur, clustering flou, clustering avec recouvrement.	18
V. Notions de base.	18
V.1 Centroïde.	18
V.2 Outlier.	19
V.3 Mesure de similarité.	19
V.4 La distance intra-clusters, la distance inter-clusters et la variance d'un cluster.	21
VI. L'algorithme de clustering k-moyennes (k-means).	21
VI.1 Présentation générale de l'algorithme k-means.	21
VI.2 L'algorithme des k-moyennes (k-means.)	22
VI.3 Détails des étapes de l'algorithme.	22
VI.3.1 Sélection d'une partition initiale.	22
VI.3.2 Mettre à jour la partition.	23
VI.3.3 Convergence.	23

VI.3.4 Complexité.	24
VI.4 Discussion sur l'algorithme k-means.	24
VII. Conclusion.	25

Chapitre 3 : Le calcul multi-partie sécurisé : Recueil ad hoc

I. Introduction.	29
II. La sécurité dans le calcul multi-partie.	29
II. 1 Les propriétés de sécurité dans le calcul multi-partie sécurisé.	30
II. 2 Le paradigme de la simulation idéal/réel.	31
II. 3 La puissance de l'adversaire.	32
III. La faisabilité du calcul multi-partie sécurisé.	33
IV. Le calcul multi-partie sécurisé.	34
IV.1 Notations d'écriture.	35
IV.2 Algorithme probabiliste à temps polynomial.	35
IV.3 Indistinguabilité calculable.	36
IV.4 Le calcul biparti.	37
IV.5 Le modèle semi-honnête.	37
IV.6 Le protocole biparti sécurisé.	37
V. L'évaluation sécurisée d'un circuit.	38
V.1 Construction d'un circuit brouillé.	39
V.2 Interaction entre deux parties.	41
V.3 Calcul de la sortie.	41
V.4 Discussion.	41
VI. Le chiffrement homomorphe.	42
VI.1 La sécurité sémantique.	42
VI.2 Propriété d'homomorphisme.	43
VI.3 Quelques schémas de chiffrement homomorphe.	44
VI.4 Le cryptosystème de Paillier.	44
VI.4.1 Notations.	45
VI.4.2 Résiduosit� composite de degré n .	45
VI.4.3 Classe de résiduosit�.	46
VI.4.4 Chiffrement probabiliste de Paillier.	47
VII. Le produit scalaire sécurisé.	48
VII.1 D�finition d'un protocole de produit scalaire.	49
VII.2 D�finition d'un protocole partag� de produit scalaire.	49
VII.3 D�finition du protocole de produit scalaire priv�.	49
VII.4 Le protocole priv� cryptographique SSP.	49
VII.5 Consid�rations pratiques.	50
VII.6 Le co�t de communication.	50
VIII. Autres primitives de s�curit�.	50
VIII.1 Les sch�mas de secret partag� additif (additive secret sharing schemes).	51
VIII.2 Les parts al�atoires.	51
VIII.3 La somme s�curis�e.	51
IX. Conclusion.	52

Chapitre 4: Préservation de privacy dans l'algorithme k-means: Etude et Analyse	
I. Introduction.	54
II. L'algorithme de clustering k-means sur un ensemble de données singulier.	54
III. Approche 1 : L'algorithme de clustering k-means sur un ensemble de données réparti verticalement.	56
III.1 Positionnement du problème .	56
III.2 Etude des travaux de préservation de privacy dans k-means sur des données distribuées verticalement.	57
III.2.1 Etude de l'algorithme de J. Vaidya et C. Clifton.	57
III.2.2 Etude de l'algorithme de S. Samet et A. Miri.	62
III.2.3 Etude de l'algorithme de M.C Doganay et T.B. Perderson.	63
III.3. Analyse et critiques.	65
IV. Approche 2 : L'algorithme de clustering k-means sur un ensemble de données réparti horizontalement.	65
IV.1 Positionnement du problème.	65
IV.2 Etude des travaux de préservation de privacy dans k-means sur des données distribuées horizontalement.	67
IV.2.1 L'algorithme de S. Jha et L. Kruger.	67
IV.2.2 L'algorithme de S. Samet et A. Miri.	70
IV.3 Analyse et critiques.	71
V. Approche 3 : L'algorithme de clustering k-means sur un ensemble de données réparti arbitrairement.	
V.1 Positionnement du problème.	71
V.2 Etude des travaux de préservation de privacy dans k-means sur des données distribuées arbitrairement.	72
V.2.1 L'algorithme de G. Jagannathan et R.N. Wright.	72
V.2.2 L'algorithme de C. Su et F. Bao.	75
V.2.3 L'algorithme de P. Bunn et R. Ostrovsky.	76
V.2.4 L'algorithme de J. Sakuma et S. Kobayashi.	79
V.3 Analyse et critiques.	80
VI. Résumé de l'analyse des de travaux de préservation de privacy dans l'algorithme k-means.	80
VII. Conclusion.	83
 Conclusion.	 84
Bibliographie.	85