

k^m -ΑΝΩΝΥΜΟΠΟΙΗΣΗ ΜΕ ΔΥΝΑΜΙΚΕΣ ΙΕΡΑΡΧΙΕΣ ΓΕΝΙΚΕΥΣΗΣ

ΠΛΗΡΟΦΟΡΙΕΣ: Όλγα Γκουντούνα, 210 772 1442, olga [at] dlablab.ece.ntua.gr

Μανώλης Τερροβίτης, 210 6990522, mter [at] imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η επέκταση και υλοποίηση του αλγορίθμου k^m -ανωνυμοποίησης δεδομένων με δυναμική δημιουργία ιεραρχιών γενίκευσης, με σκοπό την ελαχιστοποίηση της απώλειας πληροφορίας.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: C++

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Πολλοί οργανισμοί συλλέγουν δεδομένα από χρήστες, τα οποία θα μπορούσαν να δημοσιευθούν ή να διανεμηθούν για ερευνητικούς σκοπούς. Η διανομή και χρήση αυτών των δεδομένων εγκυμονεί σοβαρούς κινδύνους για την παραβίαση της ιδιωτικότητας. Ακόμα και με την απαλοιφή στοιχείων όπως το ονοματεπώνυμο, ΑΦΜ, κτλ., τα οποία προσδιορίζουν μοναδικά ένα άτομο, ο συνδυασμός άλλων στοιχείων όπως ηλικία, φύλο και ταχυδρομικός κώδικας θα μπορούσαν να λειτουργήσουν ως *ψευδο-αναγνωριστικά* και να οδηγήσουν στην ταυτοποίηση ενός ατόμου με μια εγγραφή στα δημοσιευμένα δεδομένα, και ταυτόχρονη αποκάλυψη ευαίσθητων προσωπικών στοιχείων του.

Για την προστασία της ιδιωτικότητας των ατόμων έχουν προταθεί διάφορες εγγυήσεις στη βιβλιογραφία [1, 2]. Η ικανοποίηση των εγγυήσεων διασφαλίζεται με την εφαρμογή αλγόριθμων ανωνυμοποίησης στα προς δημοσίευση δεδομένα. Πολλοί εξ' αυτών αξιοποιούν τις τεχνικές της *γενίκευσης* ή *απαλοιφής* τιμών των ψευδο-αναγνωριστικών. Σκοπός είναι κανένας συνδυασμός τιμών να μην εμφανίζεται στα δημοσιευμένα δεδομένα σε λιγότερες από k εγγραφές. Με τον όρο *γενίκευση* τιμών εννοούμε την αντικατάσταση μιας τιμής από κάποια άλλη πιο γενική (π.χ. ένα διάστημα τιμών που την περιέχει). Το σύνολο των πιθανών γενικεύσεων των τιμών κάθε ψευδο-αναγνωριστικού γνωρίσματος σχηματίζουν την ιεραρχία γενίκευσης.

Κάθε γενίκευση ή απαλοιφή τιμών ενός γνωρίσματος συνεπάγεται απώλεια πληροφορίας στα τελικά δεδομένα. Η ανωνυμοποίηση θα πρέπει να τηρεί τις εγγυήσεις για την προστασία ευαίσθητης πληροφορίας χωρίς ταυτόχρονα να καθιστά τα δεδομένα άχρηστα για στατιστική μελέτη. Στην παρούσα διπλωματική, ζητείται η ανάπτυξη αλγόριθμων με δυναμική δημιουργία ιεραρχιών γενίκευσης [3] οι οποίες θα επιτυγχάνουν ελάχιστη απώλεια πληροφορίας στην εφαρμογή της k^m -ανωνυμοποίησης [2].

ΣΧΕΤΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ:

[1] L. Sweeney. "*k*-Anonymity: A Model for Protecting Privacy." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 5, pp. 557-570, 2002.

[2] M. Terrovitis, N. Mamoulis, and P. Kalnis. "*Privacy-preserving anonymization of set-valued data.*" in Proc. of VLDB, 2008.

[3] M. E. Nergiz and C. Clifton. "*Thoughts on k-Anonymization.*" in DKE, vol. 63, no.3, p. 622-645, Elsevier Science, 2007.