Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Ηλεκτρολογων Μηχανικων & Μηχανικων Υπολογιστων

Σημειώσεις Διαλέξεων

Στοιχεία Θεωρίας Αριθμών

&

Εφαρμογές στην Κρυπτογραφία

Επιμέλεια σημειώσεων:
Ελενη Μπακαλη
Αρης Παγουρτζης

Διδάσκοντες:
Σταθης Ζαχος
Αρης Παγουρτζης

7 Νοεμβρίου 2011

1 Κλασικά κρυπτοσυστήματα

- Μονοαλφαβητικά: κάθε γράμμα του αρχικού κειμένου κωδικοποιείται με το ίδιο γράμμα πάντοτε (γενικότερα με τον ίδιο τρόπο). Περιλαμβάνονται τα κρυπτοσυστήματα: ολίσθησης (shift cipher: γενίκευση του κρυπτοσυστήματος Καίσαρα), παραλλαγή shift cipher με χρήση λέξης-κλειδί, αντικατάστασης (substitution cipher), PLAYFAIR, (affine cipher).
- Πολυαλφαβητικά: κάθε γράμμα του αρχικού κειμένου μπορεί να κωδικοποιείται με διαφορετικό τρόπο σε διαφορετικά σημεία του κειμένου. Περιλαμβάνονται τα κρυπτοσυστήματα Vigenère, AUTOCLAVE, Hill, permutation, Vernam (one-time pad), block ciphers, stream ciphers.

Λεπτομέρειες για τα παραπάνω συστήματα: [Zac07, κεφ.1] και [Sti06, ch.1].

2 Κρυπτανάλυση του κρυπτοσυστήματος Vigenère

Ορισμός του κρυπτοσυστήματος.

$K = (k_0, k_1, ..., k_{m-1})$: κλειδί
$X = (x_0, x_1, ..., x_{n-1})$: αρχικό κείμενο (plaintext)
$C = (c_0, c_1, ..., c_{n-1})$: κρυπτοκείμενο (ciphertext)
$c_i = E_K(x_i) = (x_i + k_{i \bmod m}) \bmod 26, 0 \leq i \leq n-1$: κρυπτογράφηση
$x_i = D_K(c_i) = (c_i - k_{i \bmod m}) \bmod 26, 0 \leq i \leq n-1$: αποκρυπτογράφηση

Κρυπτανάλυση.

Η κρυπτανάλυση συνίσταται στην εύρεση του μήκους του κλειδιού πρώτα και κατόπιν στην εύρεση του ίδιου του κλειδιού.

1. Εύρεση μήκους κλειδιού. Αυτό μπορεί να γίνει με δύο τρόπους:
   - Kasiski test: εύρεση patterns που επαναλαμβάνονται, πιθανή περίοδος: ΜΚΔ των αποστάσεων μεταξύ επαναλαμβανόμενων patterns. Στηρίζεται στο ότι ίδιες λέξεις του αρχικού κειμένου που η απόστασή τους είναι πολλαπλάσιο του m (μήκος κλειδιού), κωδικοποιούνται με τον ίδιο τρόπο.
   - Index of Coincidence (δείκτης σύμπτωσης): εκφράζει την πιθανότητα δύο τυχαίοι χαρακτήρες ενός κειμένου να ταυτίζονται.
     Σε δοσμένο κείμενο με f_i τη συχνότητα εμφάνισης του γράμματος i (σύμβαση: θεωρούμε μόνο κεφαλαία αγγλικά γράμματα, χωρίς κενά,

τα οποία αντιστοιχίζουμε σε αριθμούς από 0 έως 25):

$$IC(X) = \sum_{i=0}^{25} \frac{\binom{f_i}{2}}{\binom{n}{2}} = \sum_{i=0}^{25} \frac{f_i(f_i-1)}{n(n-1)}$$

Σημαντική ιδιότητα: η τιμή του δείκτη σύμπτωσης θα παραμείνει ίδια αν κάνουμε shift τους χαρακτήρες κατά k (γενικότερα κάτω από οποιαδήποτε μετάθεση).
Σε άγνωστο κείμενο αγγλικής $E[IC(X)] \cong \sum_{i=0}^{25} p_i^2 \cong 0.065$, όπου p_i η στατιστική συχνότητα του γράμματος i στην αγγλική.
Σε εντελώς τυχαίο κείμενο με αγγλικούς χαρακτήρες:

$$E[IC(X)] \cong \sum_{i=0}^{25} (\frac{1}{26})^2 = \frac{1}{26} \cong 0.038.$$

Άρα μπορούμε με μεγάλη πιθανότητα να ξεχωρίσουμε ένα τυχαίο κείμενο με αγγλικούς χαρακτήρες από ένα κανονικό αγγλικό κείμενο.
Κάνουμε δοκιμές για να βρούμε το μήκος m του κλειδιού. Δοκιμάζουμε διαδοχικά τις τιμές $m = 1, 2, \ldots$. Χωρίζουμε το κρυπτοκείμενο σε m στήλες, όπου κάθε στήλη i περιλαμβάνει τα γράμματα που βρίσκονται στις θέσεις $i + jm, 0 \leq j \leq \lceil n/m \rceil - 1$, και παίρνουμε το IC της κάθε στήλης. Αν έχουμε πετύχει το σωστό μήκος κλειδιού τότε πιθανότατα κάθε στήλη θα έχει IC αρκετά κοντά στο 0.065 (αλλιώς όλες οι στήλες θα έχουν λίγο-πολύ συμπεριφορά 'τυχαίου' κειμένου και άρα IC κοντά στο 0.038 – στην πράξη μπορεί να είναι λίγο μεγαλύτερο αν το κείμενο δεν είναι πολύ μεγάλο, συνήθως όμως είναι < 0.050).

2. Εύρεση κλειδιού. Και αυτό το βήμα μπορεί να γίνει με δύο τρόπους:

   - 1ος τρόπος: στατιστική κρυπτανάλυση στις στήλες με βάση τη συχνότητα εμφάνισης των γραμμάτων, διγραμμάτων, κ.λπ. της αγγλικής (ή γενικότερα της γλώσσας του αρχικού κειμένου).
   - 2ος τρόπος: πάλι με χρήση IC βρίσκουμε τα σχετικά shifts μεταξύ δύο συνεχόμενων στηλών. Δοκιμάζουμε με διαδοχικά shifts της μιας στήλης ως εξής: κάνουμε shift στους δείκτες των παρατηρημένων συχνοτήτων της στήλης αυτής (έστω ότι είναι η στήλη 2 και οι μη ολισθημένες συχνότητες συμβολίζονται με $f^{(2)}$): $f_i^{(2)'} = f^{(2)}_{(i+1) \bmod 26}$, και παίρνουμε τον *δείκτη αμοιβαίας σύμπτωσης*:

$$IC_m(X) = \sum_{i=0}^{25} \frac{f_i^{(1)} f_{(i)}^{(2)'}}{n^2}$$

Ο δείκτης αυτός αντιστοιχεί στην πιθανότητα δύο τυχαίοι χαρακτήρες από δύο κείμενα να ταυτίζονται, και έχει παρόμοιες ιδιότητες με τον δείκτη σύμπτωσης ως προς το ότι παρατηρώντας την τιμή του μπορούμε να συμπεράνουμε με μεγάλη πιθανότητα αν τα κείμενα είναι κανονικά αγγλικά κείμενα. Έτσι, θα έχουμε πετύχει το σωστό σχετικό shift αν το IC_m των δύο στηλών είναι 'κοντά' στο 0.065.

Έχοντας όλα τα σχετικά shift, τα πιθανά κλειδιά είναι 26, οπότε με 26 δοκιμές βρίσκουμε το σωστό κλειδί.

Η μέθοδος αυτή περιγράφεται με μικρές παραλλαγές στο [Sti06, ch.1].

3 Συστήματα Δημοσίου κλειδιού

Το κρυπτοσύστημα Σακιδίου Merkle-Hellman

Το κρυπτοσύστημα στηρίζεται σε μια ειδική περίπτωση του προβλήματος του Σακιδίου, που είναι γνωστό ως πρόβλημα Αθροίσματος Υποσυνόλων (Subset Sum).

Πρόβλημα Subset Sum.
Δίνεται σύνολο $A = \{a_1, ...a_n\} \subseteq \mathbb{N}$, και $k \in \mathbb{N}$.
Ζητείται, αν υπάρχει, $A' \subseteq A$ τ.ω. $\sum_{a_i \in A'} a_i = k$.

Το πρόβλημα Subset Sum είναι NP-πλήρες, θεωρείται επομένως υπολογιστικά δύσκολο και σε αυτό το γεγονός φαινόταν αρχικά ότι θα μπορούσε να στηριχθεί η ασφάλεια του συστήματος (παρ'όλα αυτά το σύστημα δεν είναι ασφαλές, όπως θα δούμε παρακάτω).

Περιγραφή του κρυπτοσυστήματος:
Υπεραυξητικό σύνολο (superincreasing set) λέγεται ένα σύνολο ταξινομημένο όπου κάθε στοιχείο είναι μεγαλύτερο από το άθροισμα όλων των προηγούμενων. πχ. $A = \{3, 7, 12, 25, 100, 211, 430\}$. Για τέτοια στιγμιότυπα το πρόβλημα Subset Sum ανήκει στην κλάση πολυπλοκότητας Π (άσκηση: σκεφτείτε γιατί).

Το κρυπτογράφημα μιας δυαδικής ακολουθίας $b_1...b_m$ μήκους όσο το μέγεθος του A, προκύπτει από το άθροισμα $\sum b_i ai$. Π.χ. για το παραπάνω σύνολο, $Enc(0100110) = 7 + 100 + 211 = 381$.

Ο Bob χρησιμοποιεί ως ιδιωτικό κλειδί ένα υπεραυξητικό σύνολο A, το οποίο 'καμουφλάρει' σε A' ώστε να φαίνεται στον υπόλοιπο κόσμο σαν τυχαίο, χρησιμοποιώντας m, t τέτοια ώστε $m > \sum a_i, gcd(t, m) = 1$:

$$A' = \{a'_i \mid a'_i = t \cdot a_i \bmod m\}$$

Το A' είναι το δημόσιο κλειδί του Βοβ.

πχ. $A = \{1, 3, 5, 11\}, m = 23, t = 7$ ιδιωτικά, ιδιωτικό κλειδί $t^{-1} \bmod m = 10$. (υπάρχει το t^{-1} επειδή $gcd(m, t) = 1$).

$A' = 7 \cdot A \bmod 23 = \{7, 21, 12, 8\}$ δημόσιο κλειδί.

$Enc_{A'}(0110) = 33$.

$Dec_{t^{-1},A}(33)$ βρίσκεται ως εξής: $t^{-1} \cdot 33 = 10 \cdot 33 = 330 \equiv 8 \pmod 2)3$ που στο υπεραυξητικό A προκύπτει από το άθροισμα $3 + 5$, άρα το κείμενο είναι 0110.

Επίθεση Shamir. Αν μπορούμε να βρούμε t', m' τ.ώ. το $A'' = (t')^{-1} \cdot A \bmod m'$ να είναι υπεραυξητικό τότε η αποκρυπτογράφηση $Dec_{(t')^{-1},A''}$ θα δώσει το ίδιο αποτέλεσμα με την $Dec_{t^{-1},A}$!

Παράδειγμα: για $t' = 7, m' = 15$, έχουμε $(t')^{-1} \equiv 13 \pmod 1)5$, και $A'' = 13 \cdot A \bmod 15 = \{1, 3, 6, 14\}$ είναι υπεραυξητικό.

Άσκηση: επαληθεύστε ότι η αποκρυπτογράφηση θα δώσει το σωστό αποτέλεσμα στο παραπάνω παράδειγμα, αλλά και στη γενική περίπτωση (δηλ. υποθέτοντας ότι έχουμε t', m' όπως παραπάνω ώστε $A'' = (t')^{-1} \cdot A \bmod m'$ υπεραυξητικό).

Ο Shamir [Sha84] έδειξε ότι αυτή η επίθεση μπορεί να γίνει γρήγορα.

Βιβλιογραφία

1. [Zac07]: Σημειώσεις Ζάχου, ΕΜΠ, 2007.
2. [Sti06]: D. Stinson: Cryptography: Theory and Practice, 3rd edition, CRC Press, 2005.
3. [Sha84]: Shamir, Adi (1984). "A polynomial-time algorithm for breaking the basic Merkle - Hellman cryptosystem". Information Theory, IEEE Transactions on 30 (5): 699-704.