Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

Σημειώσεις Διαλέξεων

Στοιχεία Θεωρίας Αριθμών

&

Εφαρμογές στην Κρυπτογραφία

Επιμέλεια σημειώσεων:
Ελένη Μπακάλη
Άρης Παγουρτζής

Διδάσκοντες:
Στάθης Ζάχος
Άρης Παγουρτζής

8 Οκτωβρίου 2012

1 Επισκόπηση Κρυπτολογίας

Ο όρος *Κρυπτολογία (Cryptology)* συμπεριλαμβάνει την *Κρυπτογραφία (Cryptography)*, την *Κρυπτανάλυση (Cryptanalysis)* και τις εφαρμογές τους. Πολλές φορές λέμε Κρυπτογραφία, εννοώντας Κρυπτολογία.

Αντικείμενα της Κρυπτολογίας

- Μυστικότητα / Ιδιωτικότητα (Secrecy / Privacy)
  - Κρυπτογράφηση: μετασχηματισμός *απλού αρχικού κειμένου (plaintext)* σε *κρυπτοκείμενο (ciphertext)*, συνήθως με χρήση κλειδιού.
  - Αποκρυπτογράφηση: μετασχηματισμός κρυπτοκειμένου στο αρχικό κείμενο, συνήθως με χρήση κλειδιού.
  - Παραγωγή / Διανομή κλειδιού
  - Συμμετρικά κρυπτοσυστήματα / Ιδιωτικού κλειδιού (κρυπτογραφία διπλής κατεύθυνσης): μονοαλφαβητικά, πολυαλφαβητικά, τμήματος, ροής, DES, AES
  - Μη συμμετρικά κρυπτοσυστήματ / Δημοσίου κλειδιού (κρυπτογραφία μονής κατεύθυνσης): Knapsack, RSA, ElGamal, Elliptic Curves
- Αυθεντικοποίηση (Authentication)
  - Data / message origin: υπογραφές, κυρίως βασισμένα σε συστήματα δημοσίου κλειδιού αλλά και Message Authentication Codes (MACs)
  - Entity / User: Identification Schemes, πρωτόκολλα ταυτοποίησης (Interactive Proofs (IP), Zero Knowledge (ZK))
- Ακεραιότητα (Integrity)
  - Συνήθως περιλαμβάνεται στην αυθεντικοποίηση
  - Hash Functions (επίσης έχουν μεγάλη χρήση στις ψηφιακές υπογραφές)
  - Συνδυασμός με αυθεντικοποίηση (MACs = keyed hash functions)
- Διαχείριση κλειδιών (Key Management)
  - Παραγωγή
  - Διανομή
  - Έμπιστη αρχή
- Πρωτόκολλα (πολλών συμμετεχόντων)
  - Broadcast
  - Consensus

  - Mental poker
  - Secure Function Evaluation (SFE), Secure Multiparty Computation (S-MPC)
  - Voting / Elections
  - Interactive Proofs / Zero Knowledge
- Μαθηματικό υπόβαθρο
  - Θεωρία αριθμών
  - Άλγεβρα (θεωρία ομάδων, πολυώνυμα)
  - Υπολογιστική πολυπλοκότητα
  - Σημαντικά προβλήματα: έλεγχος πρώτων αριθμών (primality), παραγοντοποίηση (factoring), διακριτός λογάριθμος (discrete logarithm)

2 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: [Zac12, κεφ.1] και [Sti06, ch.1].

3 Κρυπτανάλυση του κρυπτοσυστήματος Vigenère

Ορισμός του κρυπτοσυστήματος.

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

Τα κείμενα και το κλειδί αποτελούνται από κεφαλαία γράμματα της Αγγλικής γλώσσας (χωρίς κενά), τα οποία αντιστοιχίζουμε στους αριθμούς από 0 έως 25.

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 το πλήθος εμφανίσεων του γράμματος i:

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

     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:

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

     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 r του κλειδιού. Δοκιμάζουμε διαδοχικά τις τιμές $r = 1, 2, \ldots$. Χωρίζουμε το κρυπτοκείμενο σε r στήλες, όπου κάθε στήλη i περιλαμβάνει τα γράμματα που βρίσκονται στις θέσεις $i + jr, 0 \leq j \leq \lceil n/r \rceil - 1$, 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: με χρήση ενός δείκτη παρόμοιου με τον IC βρίσκουμε το σχετικό shift μεταξύ της πρώτης στήλης και της m-οστής στήλης (για $2 \leq m \leq r$).

  Για να γίνει αυτό δοκιμάζουμε διαδοχικά shifts της πρώτης στήλης ως εξής: για $0 \leq j \leq 25$ κάνουμε shift στους δείκτες των παρατηρημένων εμφανίσεων των γραμμάτων της στήλης κατά j ως εξής:

  έστω ότι το πλήθος εμφανίσεων του χαρακτήρα i στην πρώτη στήλη συμβολίζεται με $f_i^{(1)}$): τότε το πλήθος εμφανίσεων του χαρακτήρα i στην πρώτη στήλη όπου κάθε χαρακτήρας έχει ολισθήσει κατά j, δίνεται από τον τύπο:

  $$f_i^{(1)'} = f_{(i-j) \bmod 26}^{(1)}$$

  Στη συνέχεια υπολογίζουμε τον *Δείκτη Αμοιβαίας Σύμπτωσης (Index of Mutual Coincidence – IMC)*:

  $$IMC(C_1, C_m) = \sum_{i=0}^{25} \frac{f_i^{(1)'} f_{(i)}^{(m)}}{s^2}$$

  όπου s το πλήθος των χαρακτήρων των στηλών του κειμένου ($s = \lceil n/r \rceil$).

  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 (ή προέρχονται από αγγλικό κείμενο, με την ίδια ολίσθηση). Έτσι, θα έχουμε πετύχει το σωστό σχετικό shift αν το IMC των δύο στηλών είναι "κοντά" στο 0.065.

  Έχοντας τα σχετικά shift της πρώτης στήλης με τις υπόλοιπες τα πιθανά κλειδιά είναι 26, οπότε με 26 δοκιμές βρίσκουμε το σωστό κλειδί.

Η μέθοδος αυτή περιγράφεται (με κάποιες παραλλαγές) στο [Sti06,ch.1].

Βιβλιογραφία

Zac12 : Σημειώσεις Ζάχου, ΕΜΠ, 2012.

Sti06 : D. Stinson: Cryptography: Theory and Practice, 3rd edition, CRC Press, 2005.