Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Εθνικό Μετσόβιο Πολυτεχνείο
Ροή Λ, 9ο εξάμηνο

Υπολογιστική Κρυπτογραφία

Πρώτη σειρά ασκήσεων

Σπουδαστής
Παπασκαρλάτος Αλέξανδρος (Α.Μ.: 03111097)

Ημερομηνία Υποβολής: 21 Οκτωβρίου 2019

Οι κώδικες που ζητούνται βρίσκονται σε ξεχωριστά αρχεία.

Άσκηση 1

1.

Η Eve είναι πιθανό να χρησιμοποίησε τις μεθόδους που έχουμε δει και στο μάθημα: Kasiski test, αξιοποίηση δείκτη σύμπτωσης, αξιοποίηση δείκτη αμοιβαίας σύμπτωσης, υπόθεση συχνών λέξεων, κλπ. Θα αναφερθούμε σε αυτές τις μεθόδους περισσότερο στην επόμενη άσκηση όπου αποκρυπτογραφούμε ένα ciphertext (Vigenere) χωρίς να γνωρίζουμε το κλειδί.
Μάλιστα, εάν οι Alice και Bob χρησιμοποιούσαν το ίδιο κλειδί κάθε φορά για κάθε μήνυμά τους, τέτοιες τεχνικές "σπασίματος" γίνονται ακόμα ευκολότερα, καθώς οι πρώτοι χαρακτήρες του κάθε μηνύματος κωδικοποιούνται πάντα με τους ίδιους χαρακτήρες του κλειδιού.

Στην περίπτωση λοιπόν που η Eve χρησιμοποίησε μόνο τέτοιες τεχνικές, η κρυπτογράφηση του κλειδιού δεν ωφελεί σε τίποτα, καθώς οι τεχνικές αυτές δεν υποθέτουν αρχικά τίποτα για το κλειδί. Έτσι η Eve θα έβρισκε μόνο την κρυπτογραφημένη μορφή του κλειδιού, αλλά αυτό δεν της κάνει διαφορά, αυτό χρειάζεται για να αποκρυπτογραφήσει το μήνυμα.
Πχ στο επόμενο ερώτημα, δε θα έβρισκε τη λέξη "cryptography", αλλά τη λέξη "gvctxskvetlc" ως κλειδί και αυτό θα ήταν υπεραρκετό.

Ωστόσο, υπάρχει μία περίπτωση, όπου η κρυπτογράφηση του κλειδιού ίσως βοηθούσε, καθώς η πραγματικότητα είναι πολλές φορές απλούστερη. Ίσως να ακουστεί αφελές, αλλά η Eve μπορεί απλά να "μάντεψε" το κλειδί. Συγκεκριμένα, εάν το κλειδί ήταν μια φυσική λέξη (πολύ κακή τακτική) όπως "cryptography" και η Eve ξέρει πως οι Alice και Bob ασχολούνται με κρυπτογραφία, θα μπορούσε να έχει δοκιμάσει διάφορες σχετικές λέξεις και να το βρήκε "τυχαία".
Σε αυτό το γεγονός βασίζονται οι επιθέσεις λεξικού, όπου ως κλειδιά (ή password) δοκιμάζονται διάφορες παραλλαγές προβλέψιμων επιλογών. Για αυτό το λόγο, πχ το "password1234" είναι πολύ κακό password ανεξαρτήτως του πλήθους των ενδεχόμενων συνδυασμών 12 αλφαριθμητικών χαρακτήρων.
Σε αυτήν την περίπτωση, η κρυπτογράφηση του κλειδιού θα βοηθούσε εφόσον η Eve δε θα ήξερε πως το κλειδί είναι κρυπτογραφημένο. Συγκεκριμένα, δε θα βοηθούσε ακριβώς η κρυπτογράφηση αυτή καθ' αυτή, αλλά η μετατροπή του κλειδιού σε ένα φαινομενικά τυχαίο string (εφόσον η Eve δε θα ήξερε πως είναι κρυπτογραφημένο). Εάν, το κλειδί ήταν εξαρχής randomised (όπως και θα 'πρεπε), η κρυπτογράφηση δε θα πετύχαινε τίποτα (σε αυτήν την περίπτωση η Eve δε θα μπορούσε να κάνει επίθεση λεξικού όπως και να 'χει).
Πχ στο επόμενο ερώτημα, η Eve θα μπορούσε να μαντέψει το κλειδί "cryptography", αλλά όχι το κλειδί "gvctxskvetlc" (εφόσον η Eve δεν ξέρει πως το κλειδί είναι κρυπτογραφημένο).

Να σημειώσουμε, επίσης, πως αν η μορφή των μηνυμάτων είναι αυτή που φαίνεται στο δεύτερο ερώτημα, η Eve θα μπορούσε να αντλήσει ακόμα περισσότερη πληροφορία για το κείμενο και το κλειδί. Τα κεφαλαία γράμματα, τα κενά και οι στίξεις παρέχουν πολλή πληροφορία.
Πχ στο επόμενο ερώτημα, το μήνυμα της Alice ξεκινάει με “Nd Dhy”. Είναι εύκολο να υποθέσει κανείς πως αυτό αντιστοιχεί σε “Hi Bob”. Ίσως αυτή η υπόθεση να αποδειχθεί λάθος, αλλά είναι πιθανή. Συνεπώς, κάποιος μπορεί γρήγορα να συμπεράνει τα 5 πρώτα γράμματα του κλειδιού με μια ματιά. Όμοια με την υπόλοιπη πληροφορία που παρέχει η δομή του μηνύματος (μονογράμματες λέξεις, ποιες λέξεις ακολουθούν ένα κόμμα, ποιες λέξεις ξεκινούν με κεφαλαίο και άρα είναι ονόματα, κλπ).
Έτσι, η Eve μπορεί να συνδυάσει συντακτική και ενδεχόμενη σημασιολογική ανάλυση του κειμένου με τις παραπάνω μεθόδους για να βρει το κλειδί.

Ανακεφαλαιώνοντας, αν η Eve ξέρει τι κάνει, η κρυπτογράφηση του κλειδιού δε θα βοηθήσει.
Η Eve μπορεί να το σπάσει, μην κάνοντας καμία αρχική υπόθεση για τη μορφή του κλειδιού. Τέτοιες τεχνικές σπασίματος θα δούμε αναλυτικά στην άσκηση (2).
Αυτό που ίσως βοηθούσε θα ήταν η επιλογή μακρύτερου κλειδιού.

2.

Έγραψα κώδικα σε c++ ο οποίος κρυπτογραφεί τη λέξη “cryptography” με κρυπτοσύστημα του Καίσαρα (με διαφορετικό κλειδί κάθε φορά) και, στη συνέχεια, χρησιμοποιεί την προκύπτουσα λέξη ως κλειδί για την αποκρυπτογράφηση του μηνύματος της Alice (στο κρυπτοσύστημα Vigenere). Ο κώδικας βρίσκεται σε ξεχωριστό αρχείο (Crypto_ex_1_1.cpp).

Έτσι, προέκυψαν 26 διαφορετικά ενδεχόμενα αρχικά μηνύματα.
Επέλεξα το μόνο που έβγαζε νόημα.

Το κλειδί είναι το “gvctxskvetlc” και προέκυψε από τη λέξη cryptography με κρυπτογράφηση στο σύστημα του Καίσαρα με κλειδί 4.

Πρόκειται για το κάτωθι κείμενο:

Hi Bob. I think we have finally managed to win Eve, but we should think of a better way to encrypt our messages. I still want to use the Vigenere cipher. Can you think of a way to make our messages impossible to decrypt for her?

3.

Ο σχετικός κώδικας βρίσκεται στο ίδιο αρχείο με τον κώδικα του προηγούμενου ερωτήματος (Crypto_ex_1_1.cpp).
Η κρυπτογραφημένη (με τον ίδιο τρόπο) απάντηση του Bob:

O rqgawb dj tyauig pfdv wsmsgx ygvoqzomgr vndu fbkcvkx zt xzcwffq olx dqamex zgnz, lttn lojnim. Sa ax hcto vh jsuz snc oknutdwc dqizuyddeb ly yivcavo (vark kxlbpxoii ibjpzgm dgimgvv), oo nlhfnj pux hwin abej rzpzqz (ko pxluz) ojx iwxbxa zh upt fbkcvkx lpj, dp tavsomhy, puo txrko olx dcsz mxv (gb niepez v xxoq vzrzeje fgr qzko gty euqgk jsxt qxdugbgl lx ypv vztxzuilfnvrvp cty vabf mcegrg oo yabf gz vxlen dvl bfn). Olte ygt, vab cot fxnqszu tk mxwvxlmgwnx Lfo Omfp Rgy. Ohowyqik, ejgo mxv krjyeo dk mcgagwdwxo uu dv vxf'd wi zfgyngw xfn ri lsqagf tiky zpbxktvvx mmxxxnlvojp, lmsmzw, vlrooce iwdoikd.

Η απάντηση του Bob (αποκρυπτογραφημένη, μεταφρασμένη και με λίγο περισσότερη ανάλυση):

Εάν θέλουμε να μην μπορεί με τίποτα η Eve να αποκρυπτογραφήσει τα μηνύματα, δηλαδή να πετύχουμε τέλεια μυστικότητα, πρέπει να χρησιμοποιούμε κλειδιά με μήκος (τουλάχιστον) όσο το μήκος του εκάστοτε μηνύματος. Τότε, το σύστημά μας μετατρέπεται στο “άσπαστο” One Time Pad.
Επίσης, προφανώς να μη χρησιμοποιούμε πολλαπλές φορές το ίδιο κλειδί.

Επιπλέον, το κλειδί όπως είπαμε και στο ερώτημα (1), καλό θα ήταν να είναι ένα randomised string. Διαφορετικά, ίσως είναι ευάλωτο σε επιθέσεις λεξικού.
Ακόμα και αν είναι πολύ μεγάλο σε μήκος, ενδέχεται να είναι εύκολο να προβλεφθεί, εάν πχ είναι οι στίχοι από το αγαπημένο τραγούδι του Bob το οποίο ανεβάζει συχνά στο facebook…

Τέλος, όπως αναφέραμε επίσης και στο ερώτημα (1), καλό θα ήταν να εξαλειφθούν τα κεφαλαία γράμματα, οι στίξεις και τα κενά, καθώς παρέχουν πληροφορία για το μήνυμα.
Φυσικά, εφόσον πια χρησιμοποιούμε ουσιαστικά ένα One Time Pad, τέτοιου είδους πληροφορία δε θα βοηθήσει στη εύρεση του κλειδιού, αλλά ίσως η Eve μάθει κάτι που δεν ήξερε.
Πχ, ακόμα και με ένα θαυμαστικό, η Eve μπορεί να υποθέσει πως υπάρχει λόγος για ενθουσιασμό.

Άσκηση 2

Έγραψα κώδικα σε c++ μέσα από τον οποίο αποκρυπτογράφησα το δοθέν κείμενο.
Ο κώδικας βρίσκεται στο αρχείο (Crypto_ex_1_2.cpp).

Τα βήματα που ακολούθησα ήταν αυτά που παρουσιάστηκαν στο μάθημα.
Αρχικά, προκειμένου να βρω το μήκος r του κλειδιού, θεώρησα αυθαίρετα ένα r και χώρισα το κείμενο σε r στήλες. Στη συνέχεια, υπολόγισα το δείκτη σύμπτωσης (IC) της κάθε στήλης.
Επανέλαβα την παραπάνω διαδικασία για όλα τα ενδεχόμενα r από 1 εώς 19 (αυθαίρετο άνω όριο).

Στη συνέχεια, εξέτασα τους IC των στηλών για κάθε ξεχωριστό r και επέλεξα αυτόν με τους πιο "φυσικούς" δείκτες. Αναφέρουμε πως ένα αγγλικό κείμενο έχει IC κοντά στο 0,065 ενώ ένα string από τυχαίους αγγλικούς χαρακτήρες έχει IC κοντά στο 0,038.
Συνεπώς, μια στήλη που έχει ψηλό IC, είναι πολύ πιθανό να αντιστοιχεί σε πραγματικό κείμενο, ενδεχομένως shifted κατά μία σταθερά.
Έτσι, συμπέρανα πως στη συγκεκριμένη άσκηση, r = 7 (μάλλον).

Στο επόμενο βήμα, επέλεξα κάποιο k και έκανα shift τους χαρακτήρες κάθε στήλης (εκτός από την πρώτη στήλη). Στη συνέχεια, υπολόγισα το δείκτη αμοιβαίας σύμπτωσης (IMC) με την πρώτη στήλη.
Επανέλαβα τη διαδικασία για κάθε k από 0 εώς 25 (εφόσον το αγγλικό αλφάβητο έχει 26 χαρακτήρες).

Ταξινόμησα τα IMC για την κάθε στήλη ξεχωριστά. Επέλεξα για κάθε στήλη το k που έδινε το καλύτερο IMC. Οπότε βρήκα έτσι (πιθανά) σχετικά shift των στηλών σε σχέση με την πρώτη.

Έπειτα, δοκίμασα ένα shift, έστω k για την πρώτη στήλη και τα αντίστοιχα shift για τις υπόλοιπες (k + relative shift) και αποκρυπτογράφησα το κείμενο.
Επανέλαβα τη διαδικασία για κάθε k από 0 εώς 25.

Από τα 26 κείμενα που προέκυψαν, επέλεξα το μόνο στο οποίο μπορούσα να δω λέξεις, αλλά ακόμα και αυτό δεν ήταν εντελώς σωστό. Με λίγη ανάλυση (και λίγο trial and error), συμπέρανα πως μάλλον τα shift της δεύτερης και τρίτης στήλης ήταν λάθος.
Συνεπώς, δοκίμασα να αλλάξω τα relative shift αυτών των στηλών σύμφωνα με τους επόμενους μεγαλύτερους IMC και διάβασα τα κείμενα που προέκυψαν. Μόνο ένα έβγαζε νόημα.

Σημειώνω πως ίσως σωστότερη τεχνική στη γενική περίπτωση θα ήταν, στο βήμα που αποφάσιζα τα relative shifts, να συνεχίσω και να αποκρυπτογραφήσω για πολλά πιθανά relative shifts (πχ τα 4 μεγαλύτερα ως προς το IMC για κάθε στήλη) και στη συνέχεια να εξετάσω εν τέλει τα (πάμπολλα) κείμενα που θα προέκυπταν αν περιέχουν γνωστές και συχνές λέξεις της αγγλικής, όπως "the", "are" κοκ.
Τέλος, θα εξέταζα με το μάτι τα κείμενα που θα "περνούσαν" αυτά τα φίλτρα.

Λοιπόν, κατέληξα πως το κλειδί είναι η λέξη “CHAPLIN” και το αποκρυπτογραφημένο κείμενο είναι το κάτωθι:

TOTHOSEWHOCANHEARMEISAYDONOTDESPAIRTHEMISERYTHATISNOWUPONUSISBUTTHEPASSINGOFGREEDTHEBITTERNESSOFMENWHOFEARTHEWAYOFHUMANPROGRESSTHEHATEOFMENWILLPASSANDDICTATORSDIEANDTHEPOWERTHEYTOOKFROMTHEPEOPLEWILLRETURNTOTHEPEOPLEANDSOLONGASMENDIELIBERTYWILLNEVERPERISH

Αναφέρω πως πρόκειται για μέρος από τον τελικό λόγο του Τσάπλιν στην ταινία “The great dictator”.
Δεν έχει κάποιο νόημα στα πλαίσια της άσκησης ή του μαθήματος να το “καθαρογράψω”, αλλά ήθελα να το κάνω. Το κείμενο με φυσική δομή είναι το κάτωθι:

To those who can hear me, I say do not despair. The misery that is now upon us is but the passing of greed, the bitterness of men who fear the way of human progress. The hate of men will pass and dictators die. And the power they took from the people will return to the people. And so long as men die liberty will never perish.

Άσκηση 3

α.

Οφείλουμε να ξεχωρίσουμε δύο περιπτώσεις. Η πρώτη είναι ο επιτιθέμενος, ο κακός άνθρωπος, να μη γνωρίζει πως πρόκειται για τροποποιημένο Vigenere και η δεύτερη είναι πως γνωρίζει. Με τη δεύτερη περίπτωση θα ασχοληθούμε στο ερώτημα (β).

Θεωρούμε πως έχουμε αλφάβητο AB με μέγεθος |AB|. Πχ για αγγλικά, |AB| = 26.
Λοιπόν, έστω πως ο επιτιθέμενος φέρεται σα να έχει να κάνει με απλό Vigenere.
Τότε, με αυτήν την τροποποίηση του Vigenere είναι σα να δημιουργούμε ένα νέο μεγαλύτερο κλειδί.
Για παράδειγμα, αν έχουμε το κλειδί key = “ABCD” και k=2, τότε προκύπτει το κλειδί key’ = “ABCDCDEFEFGH…”.

Φυσικά, το νέο αυτό κλειδί κάποια στιγμή θα επαναλαμβάνεται. Μάλιστα, παρατηρούμε πως εάν το αρχικό κλειδί έχει μήκος r, τότε το νέο κλειδί θα έχει μήκος r’ = r · lcm(k, |AB|) / k . Άρα πολλαπλασιάζουμε το μήκος r του κλειδιού κατά ένα παράγοντα lcm(k, |AB|)/k .

Πχ εάν παίζουμε στο αγγλικό αλφάβητο και έχουμε το κλειδί key = “ABC” (με μήκος r = 3) και k=2, θα προκύψει το κλειδί
“ABCCDEEFGGHIIJKKLMMNOOPQQRSSTUUVWWXYYZA”
με μήκος r’ = 39 (= 3 · lcm(2, 26) / 2) .

Συνεπώς, προκύπτει πως επιθυμούμε να έχουμε όσο το δυνατόν μεγαλύτερο lcm σε σχέση με το k, ώστε να αργήσει όσο το δυνατόν περισσότερο να επαναληφθεί το κλειδί.
Επομένως, θέλουμε να επιλέξουμε k με όσο το δυνατόν λιγότερους κοινούς παράγοντες με το |AB|, ιδανικά θέλουμε να είναι σχετικά πρώτο. Τότε r’ = r · |AB|.
Για παράδειγμα, στα αγγλικά θα είναι:
k = 26 => r’ = r
k = 13 => r’ = 2 r
k = 8 => r’ = 13 r
k = 3 => r’ = 26 r

Λοιπόν, με κατάλληλο k και δεδομένου πως ο επιτιθέμενος δε γνωρίζει πως τροποποιήσαμε το Vigenere, δημιουργούμε κλειδί με μήκος |AB|-πλάσιο του αρχικού.

β.

Εδώ ας υποθέσουμε πως εμείς είμαστε ο επιτιθέμενος και, σε αντίθεση με την προηγούμενη περίπτωση, γνωρίζουμε τη μέθοδο που χρησιμοποιείται για την κρυπτογράφηση.

Λοιπόν, αρκεί να χρησιμοποιήσουμε τις μεθόδους για το Vigenere (όπως τις παρουσιάσαμε στην άσκηση (2) με τις εξής τροποποιήσεις:

Προκειμένου να βρούμε το μήκος του κλειδιού, δοκιμάζουμε κάποιο r και χωρίζουμε το κείμενο σε στήλες. Όμως, επίσης θεωρούμε κάποιο k, $0 \leq k \leq |AB|$ και κάνουμε shift το i-οστό στοιχείο σε κάθε στήλη κάτα $i \cdot k$ θέσεις στο αλφάβητο. Στη συνέχεια, υπολογίζουμε τον δείκτη σύμπτωσης για την κάθε στήλη και συγκρίνουμε με τον πρέπον για το συγκεκριμένο αλφάβητο.
Επαναλαμβάνουμε τη διαδικασία για κάθε δυνατό k (από 0 εώς |AB|).
Έπειτα, δοκιμάζουμε για διαφορετικό r, κοκ.
Βλέπουμε πως η δουλειά |AB|-πλασιάστηκε, αλλά δεν πάθαμε και τίποτα, ήταν αρκετά γρήγορο.

Στο τέλος αυτής τη διαδικασίας, θα έχουμε βρει ενδεχόμενα match για το r και το k.
Λοιπόν, άπαξ και έχουμε βρει το k, το σύστημα εκφυλίζεται σε απλό Vigenere.
Οπότε συνεχίζουμε κατά τα γνωστά (βρίσκουμε σχετικά shift με χρήση δεικτών αμοιβαίας σύμπτωσης, κάνουμε γραμματικές αναλύσεις, επιθέσεις λεξικού, κοκ).

Τελικά, παρατηρούμε πως αυτό το σύστημα δε βελτιώνει αξιόλογα τον απλό Vigenere εάν ο επιτιθέμενος γνωρίζει την τροποποιημένη μέθοδο.