Εισαγωγή στα Συστήματα Ασφάλειας Πληροφοριών
Πανεπιστήμιο Κρήτης
Άνοιξη 2024
Καθηγητής: Ευάγγελος Μαρκάτος

Εργασία 1

Αλγόριθμοι Κρυπτογραφίας & Αποθήκευση Κλειδιού-Τιμής

Κωνσταντίνα Παπαφραγκάκη (csdp1339)

Ανάθεση: 05/03/2024
Παράδοση: 19/03/2024

Οι **Ώρες Γραφείου** θα πραγματοποιηθούν κάθε *Δευτέρα* και *Τετάρτη* από τις 16:00 έως τις 18:00 στις αίθουσες *B208*, *B210* και *B212* ή απομακρυσμένα στο `https://uoc-gr.zoom.us/j/84560032286`.

Η εργασία υποβάλλεται μέσω του turnin στους υπολογιστές της σχολής χρησιμοποιώντας την ακόλουθη εντολή: **turnin assignment_1@hy457**.

Μέρος Α: Αλγόριθμοι Κρυπτογραφίας

Για αυτήν την ενότητα της εργασίας, θα υλοποιήσετε αλγόριθμους κρυπτογράφησης και θα προσπαθήσετε να τους αποκρυπτογραφήσετε χωρίς να γνωρίζετε το κλειδί. Θα πρέπει να δημιουργήσετε δύο αρχεία: το `cs457_crypto.h`, το οποίο θα περιλαμβάνει τις δηλώσεις των συναρτήσεων, και το `cs457_crypto.c`, όπου θα υλοποιήσετε τις συγκεκριμένες συναρτήσεις. Πρέπει να χρησιμοποιηθεί μόνο η γλώσσα προγραμματισμού C, χωρίς να βασίζεται σε εξωτερικές βιβλιοθήκες. Αν θέλετε να εξετάσετε τα αντίστοιχα παραδείγματα κρυπτανάλυσης, μπορείτε να επισκεφτείτε το Φροντιστήριο που παρέχεται στη σελίδα `https://www.csd.uoc.gr/~hy457/assignments.html`.

Θα πρέπει να υλοποιήσετε ένα demo (test file) για τις συναρτήσεις που περιγράφονται στο μέρος Α της εργασίας, παρουσιάζοντας επιτυχώς τις διαδικασίες κρυπτογράφησης και αποκρυπτογράφησης.

1. Ο αλγόριθμος one-time pad είναι μια τεχνική κρυπτογράφησης που χρησιμοποιεί ένα τυχαία δημιουργημένο κλειδί, συνήθως αναφερόμενο ως one-time pad, με μήκος τουλάχιστον ίσο με το μήκος του plaintext. Κατά την κρυπτογράφηση, σε κάθε bit ή χαρακτήρα του plaintext εφαρμόζεται η λειτουργία XOR με το αντίστοιχο bit ή χαρακτήρα του κλειδιού.

    - Υλοποιήστε τις συναρτήσεις **one_time_pad_encr** και **one_time_pad_decr**. Αυτές οι συναρτήσεις πρέπει να δέχονται το plaintext ή το ciphertext, το μήκος τους και το τυχαία δημιουργημένο κλειδί ως ορίσματα και να επιστρέφουν το αποτέλεσμα αντίστοιχα.

- Χρησιμοποιήστε μία γεννήτρια ψευδοτυχαίων αριθμών (pseudorandom number generator) για τη δημιουργία του κλειδιού, όπως το `/dev/urandom`. Καθώς το `/dev/urandom` παρέχει ένα νέο τυχαίο αριθμό κάθε φορά που διαβάζεται, είναι απαραίτητο να δημιουργηθεί ένα τυχαίο μυστικό κλειδί και να αποθηκευτεί στη μνήμη για να αποκρυπτογραφηθεί με επιτυχία το κρυπτογραφημένο μήνυμα.
- Υποθέστε ότι το plaintext αποτελείται μόνο από γράμματα ή αριθμούς.

2. Ο αλγόριθμος affine cipher είναι μια μέθοδος κρυπτογράφησης όπου κάθε γράμμα (είτε με κεφαλαία είτε με πεζά) αντιστοιχίζεται στον αριθμητικό του ισοδύναμο (`'x'` στο plaintext και `'y'` στο ciphertext), κρυπτογραφείται χρησιμοποιώντας μια απλή μαθηματική συνάρτηση και συγκεκριμένα την **(5x + 8) mod 26**, και στη συνέχεια μετατρέπεται πίσω σε γράμμα χρησιμοποιώντας τη συνάρτηση **21(y - 8) mod 26** για αποκρυπτογράφηση.

- Υλοποιήστε τις συναρτήσεις **affine_encr** και **affine_decr**. Αυτές οι συναρτήσεις πρέπει να δέχονται το plaintext ή το ciphertext και επιστρέφουν το αποτέλεσμα αντίστοιχα.
- Υποθέστε ότι το plaintext αποτελείται μόνο από γράμματα και/ή κενά, και το πρόγραμμα πρέπει να υποστηρίζει τα γράμματα σε κεφαλαία ή/και πεζά.

3. Να γραφτεί ένας decryptor για τον απλό αλγόριθμο αντικατάστασης που αποκρυπτογραφεί ένα ciphertext χωρίς να γνωρίζει το κλειδί (cipher alphabet).

- Αποκρυπτογραφήστε το εξής: 'Pfim im k pwbp pfkp fkm nwwx wxqjedpwt smixc pfw kzzixw krcajipfu kxt civwx km kx kmmicuwxp ix pfw Qaudspwj Mqiwxqw Twdkjpuwxp az pfw Sxivwjmipe az Qjwpw.'

Ο decryptor θα εκτελέσει τις ακόλουθες λειτουργίες σε κάθε επανάληψη:

(α′) Διαβάζει το ciphertext και ζητάει από το χρήστη να εισάγει μία αντιστοίχιση (cipher alphabet προς alphabet). Το εργαλείο θα εμφανίζει το plaintext που έχει αποκωδικοποιηθεί μέχρι στιγμής.

(β′) Ζητάει από το χρήστη να εισάγει μια εν μέρει αποκρυπτογραφημένη λέξη και το πρόγραμμα θα εκτυπώνει τις σύνηθες λέξεις που ταιριάζουν με το μοτίβο στο ciphertext. Για να το κάνει αυτό, θα ακολουθήσετε τα παρακάτω βήματα:

i. Πάρτε ως όρισμα ένα αρχείο κειμένου και το πρόγραμμα θα εκτυπώνει τη συχνότητα κάθε χαρακτήρα.

ii. Υπολογίστε τη συχνότητα του Αγγλικού Λεξικού (`https://github.com/dwyl/english-words`) και του ciphertext χρησιμοποιώντας τη λειτουργικότητα στο (i).

Για παράδειγμα:

Δίνοντας το ciphertext χωρίς καμία αλλαγή ως όρισμα και την αντιστοίχιση i -> c το παρακάτω plaintext θα τυπωθεί στην κονσόλα:

**** ** * **** **** *** **** ********* ****i *** ****** **i****** *** i**** ** ** ****i**** ** *** ******** ******* ********* ** *** ********** ** *****.
Enter partially decrypted word: th**
this, them

Next mapping: c -> e

**** ** * **** **** *** **** *****c*** ****i *** ****** **i****** *** i**** ** ** ****i**** ** *** ******** ******* ********* ** *** *********c ** *****.
Enter partially decrypted word: c*t
cat

Επανάλαβε την διαδικασία μέχρι να αποκωδικοποιηθεί ολόκληρο το plaintext και συμπεριλάβετε το plaintext στο αρχείο Readme.

4. Ο αλγόριθμος Trithemius cipher, γνωστός επίσης ως Trithemian cipher ή Tabula Recta cipher, είναι ένας πολυαλφαβητικός κρυπτογράφος αντικατάστασης που χρησιμοποιεί ένα πίνακα κρυπτογράφησης, συχνά αναφερόμενος ως tabula recta, τόσο για την κρυπτογράφηση όσο και για την αποκρυπτογράφηση. Ο πίνακας (tabula recta) αποτελείται από 26 γράμματα, ταξινομημένα κατά μήκος στο επάνω μέρος κάθε στήλης, αντιπροσωπεύοντας τα γράμματα A έως Z (ή α έως ζ, ανάλογα με το αν χρησιμοποιούνται κεφαλαία ή πεζά γράμματα στον πίνακα). Επιπλέον, ο πίνακας του αλφάβητου επαναλαμβάνεται κατά μήκος στην αριστερή πλευρά κάθε γραμμής. Κατά την πρόοδο της γραμμής προς τα κάτω, τα γράμματα μετατοπίζονται κατά μία θέση προς τα δεξιά με κυκλικό τρόπο. Για παράδειγμα, όταν το Β′ μετακινείται στην πρώτη θέση, το Α′ μετακινείται στο τέλος της γραμμής. Αυτή η κυκλική μετατόπιση εφαρμόζεται σε όλο τον πίνακα.

- Υλοποιήστε τις συναρτήσεις **trithemius_encr** και **trithemius_decr**. Αυτές οι συναρτήσεις πρέπει να δέχονται το plaintext ή το ciphertext και να επιστρέφουν το αποτέλεσμα αντίστοιχα.
- Κατά τη διαδικασία κρυπτογράφησης, κάθε γράμμα του plaintext αντικαθίσταται με το αντίστοιχο γράμμα από την κατάλληλη γραμμή του πίνακα του αλφαβήτου. Η γραμμή που χρησιμοποιείται για την αντικατάσταση καθορίζεται από τη θέση του γράμματος στο plaintext και την αντίστοιχη τιμή μετατόπισης στην ακολουθία μετατόπισης.
- Η αποκρυπτογράφηση αντιστρέφει αυτή τη διαδικασία εφαρμόζοντας την αντίστροφη μετατόπιση σε κάθε γράμμα.
- Υποθέστε ότι το plaintext αποτελείται από γράμματα, σημεία στίξης και/ή κενά, και και το πρόγραμμα πρέπει να υποστηρίζει τα γράμματα σε κεφαλαία ή/και πεζά. Κατά την κρυπτογράφηση, τα κενά και τα σημεία στίξης πρέπει να παραλειφθούν, και πρέπει να προστεθούν πίσω στο παραγόμενο plaintext μετά την αποκρυπτογράφηση.

5. Ο αλγόριθμος Scytale cipher είναι μια μέθοδος κρυπτογράφησης που χρησιμοποιεί ένα απλό μηχανισμό που είναι γνωστός ως Scytale, ο οποίος αποτελείται από ένα ράβδο και ένα λωρίδα περγαμηνής. Η ράβδος είναι συνήθως ένα κυλινδρικό αντικείμενο, όπως ένα ξύλινο κομμάτι, με συγκεκριμένη διάμετρο. Η λωρίδα περγαμηνής είναι ένα μακρύ, στενό κομμάτι υλικού, όπως περγαμηνή ή δέρμα.

- Υλοποιήστε τις συναρτήσεις **scytale_encr** και **scytale_decr**. Αυτές οι συναρτήσεις πρέπει να δέχονται το plaintext ή το ciphertext, καθώς και τη διάμετρο της ράβδου, και να επιστρέφουν το αποτέλεσμα αντίστοιχα.
- Κατά την κρυπτογράφηση, η λωρίδα περγαμηνής τυλίγεται σφιχτά γύρω από τη ράβδο, και το μήνυμα στη συνέχεια γράφεται κατά μήκος της λωρίδας.
- Η αποκρυπτογράφηση περιλαμβάνει το τύλιγμα της λωρίδας περγαμηνής γύρω από μια ράβδο με την ίδια διάμετρο που χρησιμοποιήθηκε για την κρυπτογράφηση. Όταν η λωρίδα τυλίγεται γύρω από τη ράβδο, τα γράμματα ευθυγραμμίζονται κανονικά, επιτρέποντας την ανάγνωση του αρχικού μηνύματος.
- Υποθέστε ότι το plaintext αποτελείται από γράμματα, σημεία στίξης και/ή κενά, και και το πρόγραμμα πρέπει να υποστηρίζει τα γράμματα σε κεφαλαία ή/και πεζά. Κατά την κρυπτογράφηση, τα κενά και τα σημεία στίξης πρέπει να παραλειφθούν, και πρέπει να προστεθούν πίσω στο παραγόμενο plaintext μετά την αποκρυπτογράφηση.

6. Ο αλγόριθμος Rail Fence cipher αναδιατάσσει τα γράμματα στο plaintext γράφοντάς τα σε ένα μοτίβο ζιγκ-ζαγκ πέρα από έναν προκαθορισμένο αριθμό γραμμών (rails) και στη συνέχεια τα διαβάζει για να δημιουργήσει το ciphertext. Μετά από κάθε γράμμα μιας γραμμής (rail), προστίθεται ένα κενό στο ciphertext. Η διαδικασία επαναλαμβάνεται για κάθε γραμμή (rail).

- Υλοποιήστε τις συναρτήσεις **rail_fence_encr** και **rail_fence_decr**. Αυτές οι συναρτήσεις πρέπει να δέχονται το plaintext ή το ciphertext, καθώς και τον αριθμό των γραμμών (rails) στην κρυπτογράφηση, και να επιστρέφουν το αποτέλεσμα αντίστοιχα.
- Για την κρυπτογράφηση, το plaintext γράφεται διαγώνια πάνω στις γραμμές (rails), με το ciphertext να αποτελείται από τα γράμματα της γραμμής (rail) και προστίθεται ένα κενό μετά την ανάγνωση κάθε γραμμής (rail).
- Κατά τη διαδικασία αποκρυπτογράφησης, το πλήθος των γραμμών (rails) προσδιορίζεται από το ciphertext και γράφεται ξανά διαγώνια για να παραχθεί το plaintext.
- Υποθέστε ότι το plaintext αποτελείται από γράμματα, σημεία στίξης και/ή κενά, και και το πρόγραμμα πρέπει να υποστηρίζει τα γράμματα σε κεφαλαία ή/και πεζά. Κατά την κρυπτογράφηση, τα κενά και τα σημεία στίξης πρέπει να παραλειφθούν, και πρέπει να προστεθούν πίσω στο παραγόμενο plaintext μετά την αποκρυπτογράφηση.

Μέρος Β: Αποθήκευση Κλειδιού-Τιμής

7. Γράψτε ένα πρόγραμμα σε γλώσσα C που υλοποιεί μία απλή Αποθήκευση Κλειδιού-Τιμής για την ασφαλή αποθήκευση κλειδιών και των αντίστοιχων τιμών τους σε μια βάση δεδομένων. Το πρόγραμμα πρέπει να υποστηρίζει 2 λειτουργίες: **add** (προσθήκη) ενός ζεύγους (κλειδί, τιμή) στη βάση δεδομένων και ανάκτηση είτε μιας τιμής (**read**) είτε ενός εύρους τιμών (**range-read**) από ένα κλειδί ή ένα εύρος κλειδιών αντίστοιχα.

Για να εξασφαλίσετε την ασφάλεια, πρέπει να κρυπτογραφήσετε ολόκληρη τη βάση δεδομένων, περιλαμβανομένου του αρχείου και κάθε ζεύγους ξεχωριστά, χρησιμοποιώντας τον αλγόριθμο AES. Για να το πετύχετε αυτό, θα χρησιμοποιήσετε τη βιβλιοθήκη OpenSSL (`https://www.openssl.org/`), η οποία παρέχει ένα εύρος κρυπτογραφικών αλγορίθμων.

Συγκεκριμένα, ζητείτε να υλοποιήσετε τις ακόλουθες λειτουργίες:

- **Add (Προσθήκη)** ενός ζεύγους (κλειδί, τιμή) στη βάση δεδομένων:

  $ kv add -f <filename> key value

  Με αυτή την εντολή, το πρόγραμμά σας θα πρέπει να πάρει το αρχείο βάσης δεδομένων, να κρυπτογραφήσει τόσο το κλειδί όσο και την τιμή χρησιμοποιώντας το AES με λειτουργία CBC και στη συνέχεια να τα αποθηκεύσει στο αρχείο βάσης δεδομένων. Αν το αρχείο υπάρχει, θα πρέπει να προσθέσετε τη νέα εγγραφή (ζεύγος) στο τέλος του αρχείου. Αν δεν υπάρχει, θα πρέπει να δημιουργήσετε το αρχείο και στη συνέχεια να προσθέσετε τη νέα εγγραφή. Το αρχείο βάσης δεδομένων πρέπει να είναι σε μορφή CSV. Υποθέστε ότι το κλειδί και η τιμή είναι θετικοί ακέραιοι αριθμοί.

  Επιπλέον, ο χρήστης θα πρέπει να παρέχει έναν κύριο κωδικό (master password) που θα χρησιμοποιηθεί για να δημιουργηθεί το κλειδί και το IV για την κρυπτογράφηση με τον αλγόριθμο AES.

  (Υπόδειξη: Αυτή η συνάρτηση `https://www.openssl.org/docs/man3.1/man3/EVP_BytesToKey.html` μπορεί να σας βοηθήσει).

  Για παράδειγμα:

  ```
  $ kv add -f db.txt 1 3
  Enter password: pass
  ```

  Ο χρήστης θέλει να αποθηκεύσει στην Αποθήκευση Κλειδιού-Τιμής το κλειδί **1** και την τιμή **3**. Το πρόγραμμα του ζητάει τον κύριο κωδικό πρόσβασης (master password) και ο χρήστης πληκτρολογεί τη λέξη 'pass'. Η Αποθήκευση Κλειδιού-Τιμής θα χρησιμοποιήσει τη λέξη 'pass' για να δημιουργήσει το κλειδί και το IV για την κρυπτογράφηση με AES του ζεύγους (κλειδί, τιμή) και την κρυπτογράφηση του αρχείου. Μετά τη λειτουργία, το κρυπτογραφημένο αρχείο βάσης δεδομένων θα πρέπει να φαίνεται όπως παρακάτω:

db.txt

```
key,value
<encrypted key>,<encrypted value>
```

- **Read** (**Ανάγνωση**) της τιμής από το αντίστοιχο ζεύγος (κλειδί, τιμή):

  $ kv read -f <filename> key

  Με αυτή την εντολή, το πρόγραμμά σας θα πρέπει να διαβάσει το αρχείο βάσης δεδομένων, να βρει την εγγραφή που αντιστοιχεί στο κλειδί που δίνεται (αποκρυπτογραφήστε το ζεύγος για να δείτε αν το δοθέν κλειδί ταιριάζει με το αποκρυπτογραφημένο) και να εκτυπώσει το αποκρυπτογραφημένο ζεύγος στο χρήστη.

  Για παράδειγμα:

```
$ kv read -f db.txt 1
Enter password: pass
Key: 1 has value: 3
```

  Ο χρήστης θέλει να ανακτήσει την τιμή ενός συγκεκριμένου κλειδιού. Το πρόγραμμα του ζητά τον κώδικα πρόσβασης και αυτός πληκτρολογεί τη λέξη 'pass' που χρησιμοποίησε προηγουμένως. Με αυτόν τον τρόπο, το αρχείο αποκρυπτογραφείται με επιτυχία και το αποκρυπτογραφημένο ζεύγος (κλειδί, τιμή) εκτυπώνεται στην κονσόλα.

- **Range-read** (**Ανάγνωση ενός εύρους**) τιμών από το αρχείο που ανήκουν σε ζεύγη (κλειδί, τιμή) από τα οποία key1<=key<=key2:

  $ kv range-read -f <filename> key1 key2

  Με αυτή την εντολή, το πρόγραμμά σας θα πρέπει να διαβάσει το αρχείο βάσης δεδομένων και να ελέγξει ότι το key1 είναι μικρότερο ή ίσο από το key2. Στη συνέχεια, για κάθε κλειδί σε αυτό το εύρος, βρείτε την εγγραφή που αντιστοιχεί στο κλειδί (αποκρυπτογραφήστε το ζεύγος για να δείτε αν το κλειδί ταιριάζει με το αποκρυπτογραφημένο) και εκτυπώστε το αποκρυπτογραφημένο ζεύγος στο χρήστη.

  Για παράδειγμα:

```
$ kv range-read -f 1 3
Enter password: pass
Key: 1 has value: 3
```

Ο χρήστης θέλει να ανακτήσει τις τιμές ενός εύρους κλειδιών. Το πρόγραμμα του ζητά τον κώδικα πρόσβασης και αυτός πληκτρολογεί τη λέξη 'pass' που χρησιμοποίησε προηγουμένως. Με αυτόν τον τρόπο, το αρχείο αποκρυπτογραφείται με επιτυχία και το αποκρυπτογραφημένο εύρος ζευγών (κλειδί, τιμή) εκτυπώνεται στην κονσόλα με αύξουσα σειρά.

Σημειώσεις

- Πρέπει να υποβάλετε όλα τα αρχεία .c και .h που δημιουργήσατε, ένα Makefile που τα μεταγλωττίζει, ένα αρχείο Readme που εξηγεί την υλοποίησή σας ή τα μέρη που δεν έχουν υλοποιηθεί, και ένα test file που χρησιμοποιεί τις υλοποιημένες λειτουργίες.
- Εάν υλοποιήσετε περισσότερες βοηθητικές συναρτήσεις ή μακροεντολές, εξηγήστε τη λειτουργικότητά τους στο αρχείο Readme.
- Αυτή η εργασία πρέπει να υλοποιηθεί χρησιμοποιώντας τη γλώσσα προγραμματισμού C σε μηχανήματα βασισμένα σε Linux.
- Μπορείτε να χρησιμοποιήσετε τη λίστα αλληλογραφίας του μαθήματος ή τις Ώρες Γραφείου για ερωτήσεις. Ωστόσο, διαβάστε πρώτα τα προηγούμενα email διότι η ερώτησή σας μπορεί να έχει ήδη απαντηθεί.
- Μην στέλνετε ιδιωτικά μηνύματα με ερωτήσεις στους βοηθούς διότι άλλοι φοιτητές μπορεί να έχουν την ίδια ερώτηση και όλοι αξίζουν να λάβουν απάντηση.
- Μην αποστέλλετε αποσπάσματα κώδικα ή τμήματα της υλοποίησής σας στη λίστα αλληλογραφίας όταν κάνετε μια ερώτηση.
- Ο υποβληθέν κώδικας θα ελεγχθεί για λογοκλοπή με χρήση συγκεκριμένου λογισμικού ανίχνευσης λογοκλοπών.