BGU CS / Information Security 2017-2018 Fall
Assignment 1: Vigenère Cryptanalysis

Deadline: Sunday, 12/11/2017

תרגיל זה עוסק בהצפנת Vigenère והפיצוח שלה.

ניתן לעשות את התרגיל בזוגות, ויש להגיש קובץ ZIP אחד הכולל קוד באחת משפות התכנות המקובלות (עם הערות כמקובל), וקובץ PDF עם הסברים לבודק (בתוך אותו קובץ ה-ZIP). תשובות לשאלות בחלק א' יש להקליד במחשב ולצרף כקובץ PDF (בתוך אותו קובץ ZIP). יש לענות בקצרה ולעניין.

מבוא

צופן Vigenère דומה לצופן Caesar, פרט לעובדה כי קיימים מספר מפתחות. מכיוון שמשתמשים ביותר מאשר קבוצת הצבה אחת, צופן זה נקרא צופן פולי-אלפביתי. על מנת להצפין בצורה קלה יותר, במיוחד במחשב, נקנה לכל אות ערך מספרי: A=0, B=1, ..., Z=25.

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
13	14	15	16	17	18	19	20	21	22	23	24	25

קידוד של אותיות יתבצע בצורה הבאה:

$$C_i = \left(M_i + K_{i \bmod |K|}\right) \bmod 26$$

כאשר M_i היא אות מספר i בהודעה, K_i היא אות מספר i במפתח, ו- C_i היא האות ה-i המתקבלת בצופן. לדוגמא:

Plaintext (M)	DEFENDTHE**E**ASTWALLOFTHECASTLE
Repeating key (K)	FORTIFYFO**R**TIFYFORTIFYFORTIFY
Ciphertext (C)	ISWXVIRMS**V**TAYUFZCHNYFJQRLBQC

עבור $i = 9$, נקבל:

$$C_i = \left(M_i + K_{i \bmod |K|}\right) \bmod 26 = (M_9 + K_2) \bmod 26 = (\mathrm{E} + \mathrm{R}) \bmod 26 = \mathrm{V}$$

תהליך הפענוח הפוך להצפנה:

$$C_i = \left(M_i - K_{i \bmod |K|}\right) \bmod 26$$

חלק א'

1. נניח מגדירים הצפנה Triple-Vigenère, ז"א $E_{k_1}\left(E_{k_2}\left(E_{k_3}(x)\right)\right)$ כאשר k_1, k_2, k_3 הם מפתחות באורך m אותיות, x זה ה-plaintext ו-E_{k_i} מציין הצפנה עם מפתח k_i. האם הצפנה זו יותר מאובטחת מאשר Vigenère המקורי? הסבירו.
2. האם התשובה לשאלה הקודמת תשתנה אם המפתחות k_1, k_2, k_3 יהיו באורכים שונים? הסבירו.

חלק ב'

משימה 1

כתבו פונקציות הצפנה ופענוח של צופן Vigenère.

משימה 2

שיטה חשובה בפריצה של צפנים היא חישוב מספר הופעות יחסי של כל אות מתוך ה-א"ב הכולל.

כתבו פונקציה אשר מחשבת את תדירות ההופעות של כל אות בצופן (ciphertext) כלשהו המורכב מאותיות A–Z בלבד.

משימה 3

Index of Coincidence (IC) הינו מדד לחישוב הסתברות של בחירה אקראית של אותה אות פעמיים בטקסט נתון. ניתן באמצעות מדד זה לבדוק עד כמה התפלגות של אותיות בטקסט מסוים קרובה להתפלגות אחידה. שפה שבה ההסתברות להופעה של כל אות היא אחידה, תהיה בעלת IC השווה ל-1. בשפה האנגלית IC שווה ל-1.73.

חישוב של IC עבור טקסט כלשהו נתון ע"י:

$$\mathrm{IC} = \frac{\sum_{\alpha} n_{\alpha}(n_{\alpha}-1)}{N(N-1)/\mathrm{C}}$$

כאשר n_{α} הוא מספר ההופעות של כל אות α, C הוא מספר האותיות בא"ב, ו-N הוא מספר האותיות בטקסט.
נשים לב כי ערך ה-IC נשמר עבור אוסף אותיות plaintext אשר מופו ל-ciphertext באותה צורה (עם אותו K_i).

חישוב זה על הצופן, במידה ויחזיר לנו ערך הקרוב ל-1.73, יצביע על כך שהטקסט הוצפן עם מילת קוד באורך 1. סביר להניח, כי ברוב המקרים זה לא כך, ויש לבדוק ערך IC לאורכים שונים של מילת קוד. אוסף אותיות שמופרדות במרחק קבוע ו/או במכפלה של מרחק זה, צריכים להיות בעלי IC קרוב ל-1.73.

לכן, עבור ניחוש של $|K| = 3$ — ניקח כל אות המופיעה באינדקס אשר שארית החלוקה ב-3 היא 0. נחשב IC עבור קבוצת אותיות זו. לאחר מכן, נחזור על החישוב עבור כל אות המופיעה באינדקס אשר שארית החלוקה ב-3 היא 1. נחשב IC עבור קבוצת אותיות זו, וכנ"ל עבור שארית 2. בסופו של דבר, נחשב ממוצע של כל ה-IC-ים שקיבלנו, והתוצאה היא מדד IC עבור גודל מפתח 3.

נחזור על תהליך זה עבור מרחקים שונים עד 15.

ערך IC ממוצע אשר יהיה הכי קרוב ל-1.73 יצביע על כך שהמרחק המתאים ל-IC זה שווה לאורך מילת הקוד של הצופן. כמובן, יש לקחת בחשבון כי נקבל ערכים דומים גם עבור כפולות של המרחק הנכון.

כתבו פונקציה אשר מחשבת את ערכי ה-IC עבור מרחקים מ-1 עד 15 בצופן נתון, ומחזירה את אורך המפתח הצפוי בהצפנת Vigenère.

משימה 4

כעת נחלק את הצופן לקבוצות בגודל אותו חישבנו במשימה הקודמת. נרשום כל קבוצה אחת מתחת לשנייה, כך שנקבל עמודות אשר כל עמודה מתאימה לאותיות טקסט המוצפנות ע"י אות מסוימת של מילת המפתח.

לדוגמא, עבור הצופן:

```
QPWKALVRXCQZIKGRBPFAEOMFLJMSDZVDHXCXJYEBIMTRQWNMEA
IZRVKCVKVLXNEICFZPZCZZHKMLVZVZIZRRQWDKECHOSNYXXLSP
MYKVQXJTDCIOMEEXDQVSRXLRLKZHOV
```

נניח כי אורך המפתח שניחשנו הוא 7. נחלק לקבוצות של 7 ונקבל:

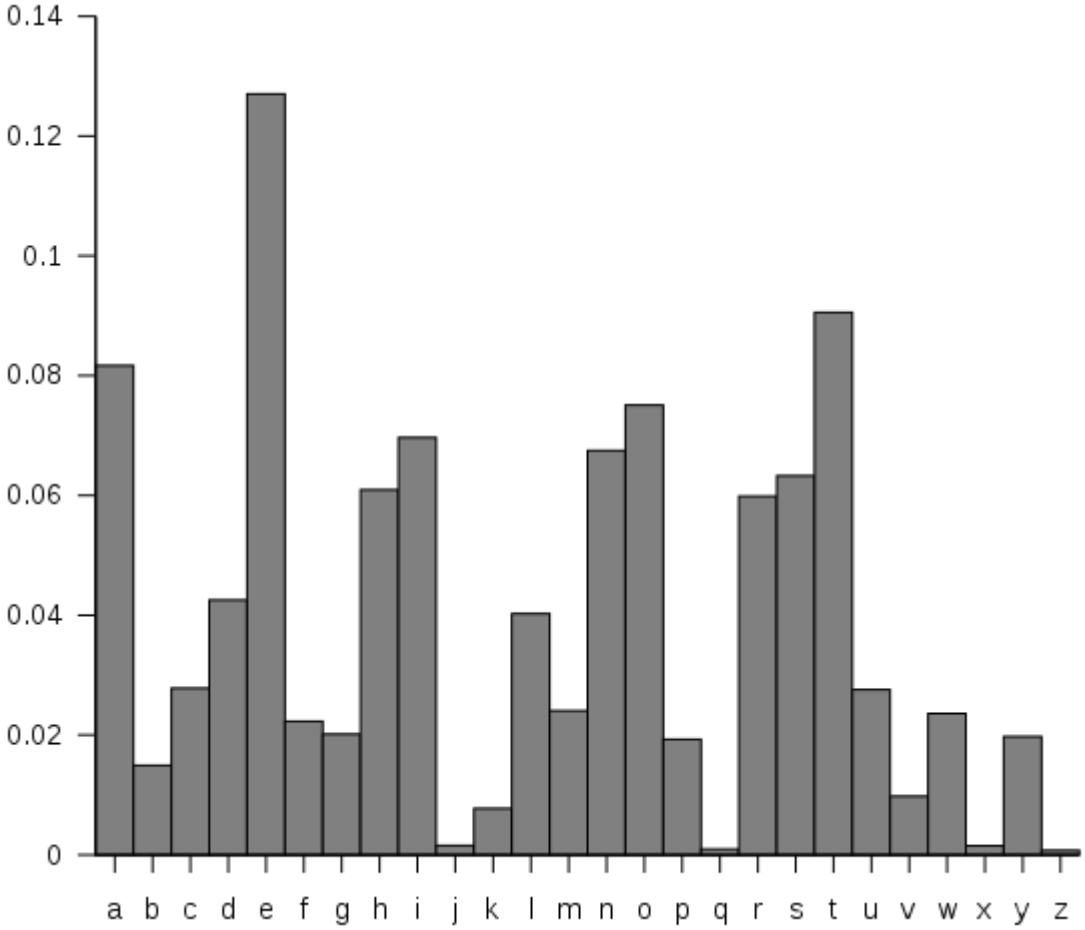
```
QPWKALV
RXCQZIK
GRBPFAE
OMFLJMS
DZVDHXC
XJYEBIM
TRQWN…
```

כל עמודה מתאימה לאות מפתח אחרת.

כדי למצוא את אותיות המפתח, נבצע את התהליך הבא **לכל עמודה**: עבור כל אות באב"ב אנגלי נחשב את השכיחות היחסית שלה בעמודה זו בצופן. לאחר מכן, עבור **כל אות מפתח אפשרית** נבצע את החישוב הבא, שמטרתו להשוות היסטוגרמת שכיחויות של כל פענוח אפשרי מול היסטוגרמה קנונית (סטנדרטית) של טקסטים באנגלית:

$$X = \sum_{\alpha} n_{\alpha} f_{\alpha}$$

כאשר n_{α} היא שכיחות מחושבת של אות מפוענחת α (יש לעבור על כל האפשרויות של אותיות המפתח), f_{α} היא שכיחות קנונית של אות α ע"פ טבלת שכיחויות ידועה מראש, ו-X הוא מדד לקורלציה בין שכיחות הופעה של אותיות בעמודה מפוענחת לבין שכיחות הופעת אותיות בטקסטים באנגלית. ככל שערכו של X גדול יותר, מידת ההתאמה טובה יותר (שכיחויות גבוהות תורמות הרבה יותר לסכום הסופי כאשר הן תואמות).



נחזור על תהליך זה עבור כל אות מפתח אפשרית מא"ב אנגלי לכל עמודה. הקורלציה הגבוהה ביותר לאות מסוימת תצביע על כך שאות זו, סביר להניח, הינה חלק ממילת מפתח עבור עמודה זו. כאשר נמצא את כל האותיות שמרכיבות את מילת המפתח, נוכל לפענח חזרה את הצופן.

כתבו פונקציה אשר מוצאת את מילת המפתח לפי התהליך שתואר. מומלץ לתכנת בצורה מודולרית, בהתאם לכל שלב בתהליך.

משימה 5

כתבו פונקציה אשר מקבלת טקסט מוצפן (ללא מילת המפתח), והיא מחזירה את הטקסט המפוענח.

פענחו את ההודעה הבאה:

```
HUGVUKSATTMUNDKUMKVVAYVLPOMCEDTBGKIIEYARTREEDRINKFSMEMQNGFEH
UVMAMHRUCPVVHBWMOGYZXVJWOMKBMAIELJVRPOMCEDRBWKIUNZEEEFRRPKMA
ZZYUDZRYRALVRZGNFLEKAKTVGNEJOAWBFLSEEBIAMSCIAKTVGNVRPKMAZHXD
YXLNFIIIDJSEMPWJOHIIBZMKOMMZNAXVRZHGTWTZNBEGFFGYAHFRKKSFRJRY
RALZSVRQGVXYIIKZHYIRHYMFMPRTTGCVKLQVMWIEBAARSDRGALFCEVOQXJID
BZVNGKIRCCWRIHVRTZHLBUKVMWIEPYSLGCXVMZKYONXHIVRKHZJYHVVVABIE
EFMNINLRWALVMJVEHDZRIIPLBOEUSJYTAAXFBJVEHDJIOHQLUVSBSNYEVLEJ
EJJFNYVFWNSEKVAWOMXUXSSJTGIAHYIWOMXUXYEIEVRQKHHZAIXZTPHVNRLB
FALVAIKREZRRMZPRGVVVNVQRELWJHZVRYVVVVZVZHYIRNYXUXZMCKZRFTKYE
CZVGTPRIUNXYBUKFFZEPAWYIPGIPNYXRIIXUKPPCEYQRYPPCEYQRPPXYFVRG
TZXZCOIEKVVJNZZRKMICTWISHYIJOOLNMUSNTJWGBSPKHZFRTAMEGJJZROIR
ROMFMVSURZTRTAMEGOMFLVQVVDWVMVVVNOVRTAMEGZRGKHRTEVXZRJLRMWIE
WVSISJQREHXVVDWVMVVVNOVRTAMEGZRGKHRTEVXZRJLRMWIEWVSITCMFBZMK
AIHAHALZNBQBKLTIENIAMSCDYNSHENVVWNXEHUKVRCIFBAEKIIKGALREOGSA
ZLVJIMWNBKMFRHEQTTXIUGCLHBVWOMKVOLRVSNMVFWPFRZFHMALVFVGGBZMN
ANRNIWMEGVRQLVKVNOPLRVYTAHIETWTZNBEAWZSWADRGEFCFUXEZXAEGPDRT
MHTGIIKNMTCTHVQOXYHFOMXUTAMJCVVPXDEJSPVRBOIRRYCBNOIIEDSCXUIU
WDHRMOIUOJVQTYOEENWGALVVAIHAHALZNBQBKLHVEKMAMVXYEYEEDUIJSKIR
KPRXLJRTBZXFOYXUXYINOIHRKPRXFZEEBUKUOPFGBUKURZEZBUKURZEZLUSD
OMXNEZIMEMHNKLHKOYVRTTFVFJVRUBXKHZWVELRTEREFNUFIOFIATUHKHZWG
BSPEENWTTCIEOOSXXUEEDOLRHUPPWJVQMOIIENTBDLRNANXUXDLZSKIEXKAF
RYPRGVVVTCMFBDLZSKIEXKEEDVRRVOSDUMQHKLHSAXOGALAFRYPRGVVVMZVR
EFXYINEAWUSKHDRTFVVVBVGXBUXFTCIPAHQSEMXHKUMEGVPYFFWFUGAVMOME
MZFHKUMEGNSBGHKRIIMUXHVUAOECIPRXSJQRMOMEGGSHWLVKHVROXMSIENYE
XSCJADHVLBVVLTXUTAMJSJQRMOMEGVXZRDMEDJAYTAXZCZPRMTIJEZXUXUAY
AOXUXYIRTDWNGKXYINQLLAIIYZBCEVVVLZXZROIRROFRLAMCLVQBFLRKAIHG
APWDYNXRKFIOPGSEXAMJTCIJBUHRNYRBMOMEGHSEXVTVNCIEXPJCUIKGALWY
UOXRKDLVNRMGATEEYVJYBYXRNYJYNAXVRDRGALVVSOICILHRSOEGXSCIAQIA
HMXYENEVGAPPDVCFHMCFRZRBMALVLZEFMVFVINEAVLQRDZLRGVXRMDRHMLWK
OKTRWVVJTVCRWOISUOAVMOQZEISSEVVUOMPNWFTVRXLRWHFFVZQLVOEDBZVQ
HVVGEMGUXKYGOIEONZXFFKEYEHWAUNXNUVZVMTGUTTFVRYSBKWIICCIQTUHJ
AOEAWUSKHDRTFVVVTCIAMOMJEWSARIMIDWITNPPZNBQLLHHWAIGLBUXFSHMY
BUKSYOLRZYEMEVRQLAIINYIPHYYDOAXUXJSLNOIATUGVIOABKLXYOPKUMOCT
RZWGULWYOMRNGKWYAQIAMOSLINEVWHVKSPVRGVGIAQIAZOEJTGCTKPQRNYEA
VPIETMEIXUARNYIEBUKWRJQGALRZGCXYRZLFRZXRESQVWCEGMOICOMHYRUED
EDWBGALVNDKUMZTCUOSABHRJHJVRJBSKHOLRKHZVNIIIXYQFRZQHVOMDAMZR
ESIUTCMFNUKRIIPLYVACTJLRTYHZSXSHKZIJOKPNBUPPTCSHZOMKSVRFPLVC
IOXYXTIRNDRTEPXKLZVRELZRNXCOHYIWOMARVHREOOLREWEXRZIVGNXYAORB
EPZZNBLHFHRSEDRTXCIIYZXJTZFCENWRWDMKHNIRBUKSIMHNUVZVHDWPAHQS
EMHBHYFZRYSEULEJTPTBGALVSXYYIAYIEYFHLAESOQIUBZGYAHFRKKSFRRMG
AZYTHIEZXHWEEQIEFVVVBPXGALVRVZRFBAXZNBPBGLPPOIXUTATCAXMQUBWK
SKSXXVRCYOLNMVRVWJVQTZMWHDWFHBPZNOLNMVRVWJVQALHZDJYGIVYINJXU
BUKWUMXUXYXYEILRNAXVRZHAHAEWEVXUXYXYEILRYSYKTZVRWAMCLDWPTYGV
LTQBKLXYAIQHMAIIEYSGALVWRDIAWZLRVZJYHDRSEASEXVRKHZQBKYSNHZAV
ESPVAQIZXHWDYCSCXZLRVZJYHDRSEASEXALVNOLRUPVUSVMQGLZVRHSEXZXR
```

ROPRWHXKHZWGBSPEENWOKVOVNWCEXWPPSJECMSCJPJORGKSLBOPRLZWRIYMJ
AHXZTPXGXYWZSDXFHUPPSOSPDHRUSOSEXJELGCXSKVQJOHIHGOEGPTQNLAII
WCSZNUQVRXMSNSHZSVWGXYJFLGSJXKJRSOEAWMSCLJARWMEJTZVGBSPYINWB
GNWFNZFHKKIEBJVRMPPCTCIQBYKVSJJUBZLFPZXUTAQVLVRPAVPPBPVQXUFF
RZSSGLZVRIIIXYQFRZFHMALVRVZRGZXZLGFRZBMCIIKNESQPFVRPRPRKONQV
EPRXSOVNBNLKIRLRXSIUAXYFAPSEEYWRTAMEFMSAMVJSIMHNGKFLSOEAWKSF
ROLRGBTFNOLROLPMEOWVGRMEGDFRMVSBMTWREMXFLDRXBUKWAIGLNUXFFVRP
RALZNFMAZDLRTOLVLVQZNJYFUPVUOACBKLAYAOXUBZKIIHYAZHMELTKUTZXC
YBEHGAEEDJQVGVYJBDVQHMCFRZQRTUXZNXVBTRMEGIIIXYQFRZXUNZMJAOIA
ZHKVDDRTNLWJIIKONARFSTPYTIPVESTEXZWZNBXBMOIWORPJAVWVFDIERLCV
SISJUBVEEYMAMVQPBJWBFZGFRZXUBZEEDHSEXPWRTYMIBUMEGRMGATCYEVHN
MLEJEMIPEPRZNBSAMOITUNLVHUWMEGZRMSMEIIKGAHXKHZPNFWPZGCXTEVEK
EYSRKIYKWCSFXCICVZXIBVPVTGMABUKNIOLGALPRMKPVZOXXLJEGBUKFEMWU
XZLRLGTEXZWRHIIIXYQFRZXUXUQVTCSHZOXKHZEVKNVVWYIALLVGEMJHFLHW
RJQNGBRJEZRPXUWVRNAHGNFPSZVNIOMDWCSFXMSFTAEYEZXZNFPRWVRKHZXH
YAIUFGSBKDVVTXLVVYMVDOLLZVHYAOLYXUXKHZIORALVSZEAZLPJHZLNMOWV
NOXUXLVVSKMGXYIJPDXRTUHEEKIAMOIWRJQGAFQVMJVVXZSWLZRBKLULAAJB
JBEWFOLVLRMEDIICXUXYEVRQYVVXEOXUBZPFSOPRGVVVQPSGAALVRVZRGUIM
EMQBKLTIOKLRMZEZDDXUBUKFFZZVEWVFPCIGLAMCLDJOBYHFRYIIBSAYEOLR
KAIDPOIELLRKOMAUXALVROIZILWKTJWFXKXYEZLRKLEJHJVRWLWFLVXRRLXR
LGYAWHYETZHBGALZSYIFXYXCAIHRGJLRNOIQHUXYINLBFLFPHJVEHYLRUIXR
WAICLHIGKBPPIDQCEVVVINXUXYIZSOLRKLFRLHMAZPPVAYXRESQVTZPYFLMZ
MKPBKLULOOLGALVRVZRAXCIIMJVRIYSGHZXFTPHZTCMAZVJVVDPCKVTYEOWG
BSPZFWMEWVVUEQMYUFXYAOLRTCIETCEGULRUSVFBOLYJBTXUTAKFDRIOHALR
DJVRMLPCTCMFLVYCWDXULVVIORPNWLRZFRMGAPRKHZHVLAEETVMQXURZTNLN
ESGCANTNLHMETZHZTPHVNRLBFALVAIKREZRRMZPRGVVVCGEFIHVRRZEAWYEU
IVRGFHMUEIAUHTXYEVRTXSWEAHIYXUSIELYBMOXYEMEIXURVVZVZHYISEOLN
MDSIDJYELPKEOATNKAMEGWMEWVVWIZRQBZLIIZORWBTJTVVGBUKXEOXUXLFR
CFMAMVXYEOIZILWKAIHGALRZGCXFISYKOIMNGZLFRZPRTCIEOWPNVRTCUHIN
LHXFKZRBYALRTGMRMOCJOPPUTALJPJORGSIRVZQLEVRVLDRRLZYEBMSXXUUL
IOXUXIYJTVFBOLQPDJSEMHOVTCCOXHOWRJQBNAQPHZEEMHRUTVORMOCWOMQS
KVQFFAQLWVSIQPSGAALVRVZRGUIMEMQBKLEEDOLRKHZVNIIIXYJCIOXVGNWK
IGPVLZMKTDRTLAMCLDWFBAXZNBSAMOIGAGPVWIYJTJJCTSPRSEYFMHFFVZQL
VOEDBZVQHVVRNYLVLLCVSCEIXHPCTCIFXLQZNBSSTKIDOIWGAHXZSYVRTTME
GVRQMOICAHTYBNLKOZVUBTWKRZEZBUKKHMSJLALVSCEQHDSETCISEVSIAIHZ
RZSLLAVBFVYKTCEGLOEUORXUTAPZENJYHHXZNBSAMOIWLJSELOECLWIYBMXV
DIIIXYQFRZ