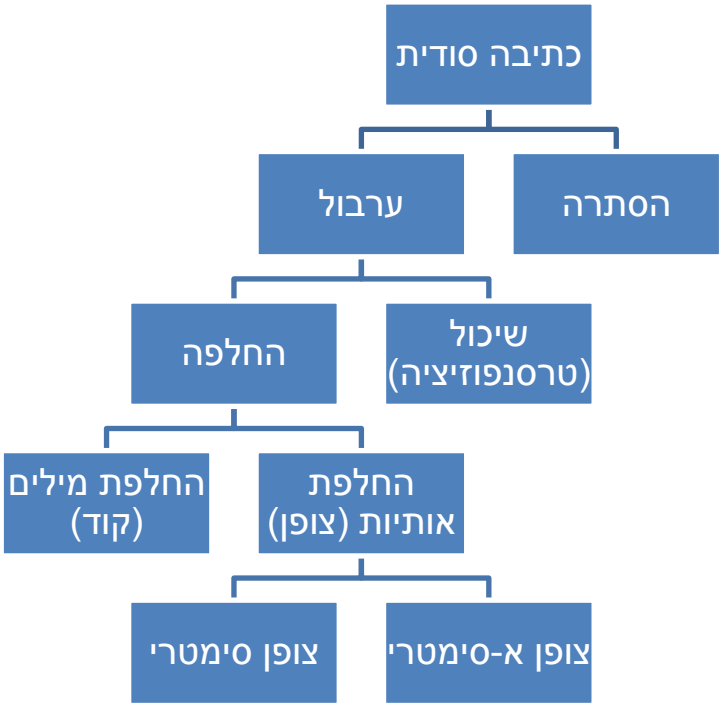
digital whisper
Technologic Papers

קריפטוגרפיה - חלק א'

מאת אופיר בק

הקדמה

ענף הכתיבה הסודית הוא גדול מאוד ובעל שורשים קדומים מאוד. אנחנו נעסוק בהצפנה באנגלית, לשם הנוחיות, אך מרבית העקרונות הם זהים לגמרי. בשביל רושם ראשוני, בחרתי לתאר את הענפים המרכזיים של הענף.



הסתרה - ניסיון להחביא את המסר עצמו, כך שלא ניתן יהיה לגלות אותו. במציאות המודרנית שלנו זה לא כל כך אפשרי, כי כשאתה מנסה להעביר מסר באינטרנט אתה עדיין חייב להשתמש בהעברה של פקטות, והן לא יהיו נסתרות.

ערבול - שינוי הטקסט שמופיע, מתוך כוונה שהוא לא יהיה ברור לאף אחד.

שיכול - שינוי סדר האותיות. דוגמה בסיסית לשיטה תהיה להפוך את סדר האותיות. השיטה של היפוך הסדר לא תהיה יעילה במיוחד, כמו בדוגמה הבאה: ?siht daer uoy nac

החלפה - החלפת האותיות באותיות אחרות. דוגמה בסיסית תהיה שימוש בצופן הקיסר, בו אנחנו קובעים מספר מסוים שישמש למספר ההיסט שלנו. לדוגמה, אם נבחר את המספר 1, האות a תהפוך לאות b, האות b ל-c וכן הלאה, כך שהאות z תהפוך ל-a, המסר 'can you see this?' יהפוך ל-'dbo zpv tff uijt?'.

צופן סימטרי - צופן שאפשר להפוך את תהליך ההצפנה ולחשוף את המסר בקלות, לדוג' הכפלה ב-2 היא תהליך שניתן להפוך בקלות, ע"י חילוק באותו המספר.

צופן א-סימטרי - לא ניתן להפוך את תהליך ההצפנה לאחור בקלות. לשם כך נהוג להשתמש בפונקציות חד כיווניות, שאין להם פענוח מוגדר בעזרת חישוב הפוך.

בחלק הזה אנחנו נעסוק בעיקר בהצפנות ישנות, ובבסיס של הפיצוח שלהן. ההצפנות האלו לא יעילות במיוחד כיום, אבל התחרות בין מפצחי הצפנים לבין מפתחי הצפנים היא הבסיס לכל תהליך ההתקדמות האנושית בנושא הקריפטוגרפיה.

שימו לב! פעמים רבות, מסירים את הרווחים מהטקסט המוצפן, כדי להקשות על חשיפתו, אך לשם ההבנה והנוחות, לא נעשה זאת.

צופן הקיסר

את הרעיון שמאחורי הצופן הזה כבר הזכרנו, אבל הפעם נרחיב קצת עליו. הצופן היה בשימוש על ידי יוליוס קיסר, וזהו מקור השם שלו. בצופן אנחנו בוחרים אות להיסט או מילה (או מספר מילים) בה נשתמש להתחלה ואחריה נמשיך בעזרת האותיות שאחרי האות האחרונה. הצופן הוא מונואלפביתי, מה שאומר שמשתמשים בסט אחד של אותיות חלופיות ביחס לאותיות המקוריות (לכל אות במסר המוצפן יש משמעות אחת בלבד - אות ספציפית במסר המקורי).

הקושי לפצח את הצופן גדל כשמשתמשים באוסף אקראי של אותיות בתור מפתח, מה שגורם לכך שיש כ-400,000,000,000,000,000,000,000,000 אופציות שונות, ומקשה אפילו על המחשב המודרני לחשוף את המסר המקורי. השיטה הראשונה לפענוח של הצופן הזה הגיעה אלפי שנים לאחר מכן, בתקופת הפריחה הערבית בתחומי המדעים. הערבים ספרו את האותיות בכמו עצומה של ספרים, והגיעו למסקנה מהו האחוז הסטטיסטי של השימוש בכל אחת מהאותיות.

מצורפת לכאן טבלת התדירות של השפה האנגלית:

אות	תדירות (%)	אות	תדירות (%)
A	8.167	N	6.749
B	1.492	O	7.507
C	2.782	P	1.929
D	4.253	Q	0.095
E	12.702	R	5.987
F	2.228	S	6.327
G	2.015	T	9.056
H	6.094	U	2.758
I	6.966	V	0.978
J	0.153	W	2.361
K	0.772	X	0.150
L	4.025	Y	1.974
M	2.406	Z	0.074

עם זאת, עדיין ישנו חסרון בשיטה הזאת, מכיוון שבמסרים קצרים התדירות לעיתים קרובות לא תהיה נכונה. כמובן שמלבד האות e, שנמצאת הרחק מהשאר, אי אפשר באמת לדרג ככה, מכיוון שעבור חלק מהאותיות ההפרשים קטנים מאוד. לכן, ההמשך של השיטה לפיצוח הצופן שונה, אך מתבססת גם היא על התדירות. אנו יודעים שלאחר האות e האות שמופיעה הכי הרבה פעמים היא האות i, ולכן, אם נאתר אות שחוזרת על עצמה הרבה מאוד פעמים לאחר האות שאנו חושדים שהיא e, אנו יכולים לחשוד שהיא i.

בנוסף, אין הרבה מילים בנות אות אחת, ולכן אם אחת מהן תופיע הרבה, ניתן לחשוד שהיא i. גם במילים בנות 3 אותיות יש סטטיסטיקה ברורה, כאשר המילים הנפוצות הן the ו- and, ואנחנו יכולים לאתר אותן, לאחר שכבר זיהינו את האות e, ועל ידי כך לאתר בבת אחת חמש אותיות נוספות.

מכאן הלאה, ניתן להוסיף עוד אותיות על ידי זיהוי מילים, ובעזרת ההיגיון לחשוף בקלות מסר שלם. לשם התרגול, אני אפתור פה מסר קצר, ואצרף מסר נוסף אותו אתם תוכלו לפתור.

בשביל הנוחות, נהוג לסמן את האותיות המוצפנות באותיות גדולות, ואת שפענחנו באותיות קטנות:

PCQ VMJYPD LBYK LYSO KBXBJXWXV BXV ZCJPO EYPD KBXBJYUXJ LBJOO KCPK. CP LBO LBCMKXPV XPV IYJKL PYDBL, QBOP KBO BXV OPVOV LBO LXRO CI SX'WJMI, KBO JCKO XPV EYKKOV LBO DJCMPV ZOICJO BYS, KXUYPD: 'DJOXL EYPD, ICJ X LBCMKXPV XPV CPO PYDBLK Y BXNO ZOOP JOACMPLYPD LC UCM LBO IXZROK CI FXKL XDOK XPV LBO RODOPVK CI XPATOPL EYPDK. SXU Y SXEO KC ZCRV XK LC AJXNO X IXNCMJ CI UCMJ SXGOKLU?'

OFYRCDMO, LXROK IJCS LBO LBCMKXPV XPV CCO PYDBLK

היות והסברנו כבר שבדיקת כל המפתחות האפשריים אינה אפשרית, אנו נעשה שימוש בניתוח תדירויות. בדיקה קצרה של הטקסט המוצפן שלנו מביאה לנו את הטבלה הבאה:

האות	מקרים	אחוזים	האות	מקרים	אחוזים
A	3	0.9	N	3	0.9
B	25	7.4	O	38	11.2
C	27	8.0	P	31	9.2
D	14	4.1	Q	2	0.6
E	5	1.5	R	6	1.8
F	2	0.6	S	7	2.1
G	1	0.3	T	0	0.0
H	0	0.0	U	6	1.8
I	11	3.3	V	18	5.3
J	18	5.3	W	1	0.3
K	26	7.7	X	34	10.1
L	25	7.4	Y	19	5.6
M	11	3.3	Z	5	1.5

האותיות שמופיעות הכי הרבה הן O, X ו-P, אך בגלל הקרבה שלהן, והסיכוי לסטייה בכמות קטנה של תווים, אנו נבדוק את סמיכות האותיות שלהן:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
O	1	9	0	3	1	1	1	0	1	4	6	0	1	2	2	8	0	4	1	0	0	3	0	1	1	2
X	0	7	0	1	1	1	1	0	2	4	6	3	0	3	1	9	0	2	4	0	3	3	2	0	0	1
P	1	0	5	6	0	0	0	0	0	1	1	2	2	0	8	0	0	0	0	0	0	11	0	9	9	0

ניתן לשים לב בקלות לכך שהאות O נמצאת בשכנות לכל אות מלבד 7, ו-X שכנה לכל אות מלבד 8. מכאן ניתן להסיק שהן כנראה תנועות. האות P לעומת זאת, מופיעה בסמיכות לאותיות ספורות בלבד, ולא מופיעה בשכנות ל-15 אותיות. דבר זה מצביע על כך שהיא עיצור.

אז האותיות X ו-O מייצגות ככל הנראה את האותיות a ו-e, שהן התנועות הנפוצות ביותר באנגלית, אך השאלה היא איזו אחת מהן היא e ואיזו אחת היא a. הרמז שיכול לעזור לנו הוא שהצירוף OO מופיע פעמיים, בזמן שהצירוף XX לא מופיע כלל.

היות והצירוף ee נפוץ יותר מאשר aa, ניתן להניח ש-O=e ו-X=a. בנוסף לכך, הטענה שלנו נתמכת על ידי שהאות X נמצאת כמילה בפני עצמה בטקסט, והאות a מייצגת את אחת משתי המילים היחידות באנגלית שמיוצגות על ידי אות אחת בלבד. האות היחידה הנוספת שמופיעה לבד בטקסט היא האות Y, ולכן סביר מאוד שהיא מייצגת את האות i, שהיא האופציה השנייה למילה שמיוצגת על ידי אות אחת בלבד. עכשיו אנו יודעים כבר ש: X=a, O=e ו- Y=i.

השלב הבא הוא שימוש רחב יותר באות e. האות e נמצאת לעיתים קרובות אחרי האות h, אך לעיתים רחוקות לפניה. לכן נספור את מספר הפעמים שהאות O מופיעה לפני אותיות אחרות ואחריהן:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
אחרי O	1	0	0	1	0	1	0	0	1	0	4	0	0	0	2	5	0	0	0	0	0	2	0	1	0	0
לפני O	0	9	0	2	1	0	1	0	0	4	2	0	1	2	2	3	0	4	1	0	0	1	0	0	1	2

ניתן לשים לב ליחס הא-סימטרי שיש לאות B עם האות O, ומכאן להסיק ש- B=h. עכשיו ניתן כבר להתחיל להשלים מילים וכך לחשוף אותיות נוספות. הטקסט שלנו הוא עכשיו:

PCQ VMJiPD LhiK LiSe KhahJaWaV haV ZCJPe EiPD KhahJiUaJ LhJee KCPK. CP Lhe LhCMKaPV aPV IiJKL PiDhL, QheP Khe haV ePVev Lhe LaRe CI Sa'aJMI, Khe JCKe aPV EiKKeV Lhe DJCMPV ZeICJe hiS, KaUiPD: 'DJeaL EiPD, ICJ a LhCMKaPV aPV CPe PoDhLK i haNe ZeeP JeACMPLiPD LC UCM Lhe IaZReK CI FaKL aDeK aPV Lhe ReDePVK CI aPAiePL EiPDK. SaU i SaEe KC ZCRV aK LC AJaNe a IaNCMJ C UCMJ SaGeKLU?

eFiRCDMe, LaReK IJCS Lhe LhCMKaPV aPV CPe PiDhLK

עכשיו ניתן להשלים מילים. המילים בנות שלוש אותיות הנפוצות ביותר באנגלית הן the ו- and. מכאן ניתן להניח ש-L=t, P=n ו- V=d, כך שהטקסט החדש הוא:

nCQ dMJinD thiK tiSe KhahJaWad haD ZCJne EinD KhahJiUaJ thJee KCnK. Cn the thCMKand and IiJKt niDht, Qhen Khe had ended the are CI Sa'aJMI, Khe JCKe and EiKKed the DJCMnd ZeICJe hiS, KaUinD: 'DJeat EinD, ICJ a thCMKand and Cne noDhtK I haNe Zeen JeACMntinD tC UCM the IaZReK CI FaKt aDeK and the ReDendK CI anAient EinDK. SaU I SaEe KC ZCRV aK tC AjaNe a IaNCMJ CI UCMJ SaGeKtU?

eFiRCDMe, taReK IJCS the thCMKand and Cne niDhtK

המילה הראשונה במשפט השני היא Cn, והיות ובכל מילה יש תנועה, C היא גם תנועה. התנועות שנותרו לנו הן u ו- o. u אינה מתאימה ולכן המילה שלנו היא on, והאות C=o. ישנה גם המילה Khe, שיכולה להיות the או she. היות ו-L=t, K=s. לאחר ההצבה הזאת יש לנו את הביטוי thoMsand and one niDhts. ניחוש הגיוני הוא שמדובר ב-thousand and one nights, ונראה כי השורה האחרונה מספרת לנו כי הקטע לקוח מ-tales from the thousand and one nights, ומכך אנו יכולים להסיק ש-R=l, D=j, J=r, I=f, M=u ו-S=m. אנו יכולים להמשיך ולחשוף מילים, ולשם כך נרשום פעם נוספת את הטקסט שברשותנו:

noQ during this time shahraWad hag Zorne Eing shahriUar three sons. on the thousand and first night, Qhen she had ended the tale of ma'aruf, she rose and Eissed the ground Zefore him, saUing: 'great Eing, for a thousand and one noghts i haNe Zeen reAounting to Uou the faZles of Fast ages and the legends of anAient Eings. maU i maEe so ZolV as to AjaNe a faNour of Uour maGestU?
teFilogue, tales from the thousand and one nights

לאחר כמה הבנות נוספות, אנו מקבלים את הצופן השלם:

a	b	c	d	e	f	G	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	רגיל
X	Z	A	V	O	I	D	B	Y	G	E	R	S	P	C	F	H	J	K	L	M	N	Q	T	U	W	מוצפן

והטקסט השלם הוא:

> Now during this time Shahrazad had borne king Shahryar three sons. On the thousand and first night, when she had ended the tale of Ma'aruf, she rose and kissed the ground before him, saying: "great king, for a thousand and one nights I have been recounting to you the fables of past ages and the legends of ancient kings. May I make so bold as to crave a favour of you majesty?"
>
> Epilogue, tales from the thousand and one nights

שימו לב, שגם את המפתח עצמות היה ניתן לגלות לקראת הסוף, ולחסוך כמה גילויים של אותיות. המפתח שנמצא כאן הוא AVOIDBYGERSPC ככל הנראה. צריך לשים לב שיש כנראה הורדה של אותיות שחוזרות על עצמן. ניחוש לא סביר, אך במקרה הזה בהחלט נכון הוא A Void by Georges Perec.

הטקסט שאתם תפצחו (אם תבחרו לנסות) הוא ארוך יותר. אני משתמש באנגלית בריטית בצפנים שלי, אז שימו לב שלעיתים האיות הוא שונה במעט (כמו favour ולא favor בטקסט הקודם):

> BT JPX RMLX PCUV AMLX ICVJP IBTWXVR CI M LMT'R PMTN, MTN YVCJX CDXV MWMBTRJ JPX AMTNGXRJBAH UQCT JPX QGMRJXV CI JPX YMGG CI JPX HBTW'R QMGMAX; MTN JPX HBTW RMY JPX QMVJ CI JPX PMTN JPMJ YVCJX. JPXT JPX HETW'R ACUTJXTMTAX YMR APMTWXN, MTN PBR JPCUWPJR JVCUFGXN PBL, RC JPMJ JPX SCBTJR CI PBR GCBTR YXVX GCCRXN, MTN PBR HTXXR RLCJX CTX MWMBTRJ MTCJPXV. JPX HBTW AVBXN MGCUN JC FVBTW BT JPX MRJVCGCWXVR, JPX APMGNXMTR, MTN JPX RCCJPRMEXVR. MTN JPX HBTW RQMHX, MTN RMBN JC JPX YBRX LXT CI FMFEGCT, YPCR CXDXV RPMGG VXMN JPBR YVBJBTW, MTN RPCY LX JPX BTJXVQVXJMJBCT JPXVXCI, RPMGG FX AGCJPXN YBJP RAMVGXJ, MTN PMDX M APMBT CI WCGN MFCUJ PBR TXAH, MTN RPMGG FX JPX JPBVN VUGXV BT JPX HBTWNCL. JPXT AMLX BT MGG JPX HBTW'R YBRX LXT; FUJ JPXE ACUGN TCJ VXMN JPX YVBJBTW, TCV LMHX HTCYT JC JPX HBTW JPX BTJXVQVXJMJBCT JPXVXCI. JPXT YMR HBTW FXGRPMOOMV WVXMJGE JVCUFGXN, MTN PBR ACUTJXTMTAX YMR APMTWXN BT PBL, MTN PBR GCVNR YXVX MRJCTBRPXN. TCY JPX KUXXT, FE VXMRCT CI JPX YCVNR CI JPX HBTW MTN PBR GCVNR, AMLX BTJC JPX FMTKUXJ PCURX; MTN JPX KUXXT RQMHX MTN RMBN, C HBTW, GBDX ICVXDXV; GXJ TCJ JPE JPCUWPJR JVCUFGX JPXX, TCV GXJ JPE ACUTJXTMTAX FX APMTWXN; JPXVX BR M LMT BT JPE HBTWNCL, BT YPCL BR JPX RQBVBJ CI JPX PCGE WCNR; MTN BT JPX NMER CI JPE IMJPXV GBWPJ MTN UTNXVRJMTNBTW MTN YBRNCL, GBHX JPX YBRNCL CI JPX WCNR, YMR ICUTN BT PBL; YPCL JPX HBTW TXFUAPMNTXOOMV JPE IMJPXV, JPX HBTW, B RME, JPE IMJPXV, LMNX LMRJXV CI JPX LMWBABMTR, MRJVCGCWXVR, APMGNXMTR, MTN RCCJPRMEXVR; ICVMRLUAP MR MT

XZAXGGXTJ RQBVBJ, MTN HTCYGXNWX, MTN UTNXVRJMTNBTW, BTJXVQVXJBTW CI NVXMLR, MTN RPCYBTW CI PMVN RXTJXTAXR, MTN NBRRCGDBTW CI NCUFJR, YXVX ICUTN BT JPX RMLX NMTBXG, YPCL JPX HBTW TMLXN FXGJXRPMOOMV; TCY GXJ NMTBXG FX AMGGXN, MTN PX YBGG RPCY JPX BTJXVQVXJMJBCT. JPX IBVRJ ACNXYCVN BR CJPXGGC.

בהצלחה! הראשון שיצליח, אדאג לציין את שמו בתור המנצח במאמר הבא בסדרת מאמרים זו.

לסיכום

דיברנו על הצפנה בסיסית, והתחלנו עם צופן הקיסר, דיברנו על איך הוא פועל והדגמנו על הפיצוח שלו. במאמר הבא נעסוק בצפנים קצת יותר מורכבים.

על המחבר

שמי אופיר בק, בן 16 מפתח תקווה. אני לומד בתכנית גבהים של מטה הסייבר הצה"לי וב-C-security, לאחר שסיימתי את לימודי המתמטיקה והאנגלית בכיתה י'. קשה למצוא חומר מעודכן בעברית, ולאחר ש-DigitalWhisper היווה עבורי מקור מידע נגיש, רציתי לתרום חזרה. ניתן ליצור איתי קשר בכתובת האימייל הבאה: ophiri99@gmail.com.

קישורים לקריאה נוספת

תדירות אותיות:

https://en.wikipedia.org/wiki/Letter_frequency

צופן קיסר:

https://en.wikipedia.org/wiki/Caesar_cipher