Homework 1

John Carlyle

January 23, 2014

1.

Symmetric cipher A symmetric cipher (single-key encryption) is a cipher such that given a plaintext P, a key K, an encryption algorithm E, and a decryption algorithm D the following is true: $P = D(E(P, K), K)$. In other words after encrypting with K the way you get the message back is to call D with the same key K.

Since the key has to be the same it is a requirement that both parties involved have access to the key. Both parties must have at some point agreed upon a key to use. Since the key is used to decrypt as well as encrypt this means that in most cases the decryption algorithm is bascially the encription algorithm in reverse. A final requirement given by the book is that the encryption algorithm has to be strong enough to withstand an opponent who has access to multiple ciphertexts that use the same key. I assume this is because the key sharing problem may mean that the same key is reused often between parties using this method to communicate. If they could exchange a key safely, why can't they just exchange the message itself through this safe channel?

2.

Linear function block cipher A choice of each bit in the 128 bits as a 1 while the rest are 0s would allow one to combine via xor to make any bit pattern. And since we know each of the individual ciphertexts plaintext equivilent we can split it since EL is a linear operator and decrypt each piece individually. We get to know the plaintext since we choose those 128 plaintexts specifically. Now they get xored together to show us the original message. Very insecure!

3.

Substitution cipher My first idea was to re-write a program I had written to break the simple substitution cipher on the cards against humanity 12 days of christmas puzzle. A program that just looked at letter frequency and a dictionary. But what fun would that be? Instead I decided to see if a genetic algorithm could crack it. Instead of just launcing into it I first looked for some research on the topic online and found this: `http://people.cs.uct.ac.za/~jkenwood/JasonBrownbridge.pdf` I stole all his constants and techniques for selection, touranments, elitism, mutation, population and crossover. Since all the aformentioned constants are hard guess well on the first try and can take awhile to tune, which isn't much fun.

Implementation I implemented the GA in python (see attached file ga-1.py), and used trigram frequencies to calculate the fitness of each key. The common english trigrams were initially collected by scanning A Tale of Two Cities by Charles Dickens. After the first attempt at decrypting the cyphertext (100 generations) I had recieved the following result as my fittest individual (key):

KEY: BDFJKHMAWTVPGULNYZCXSIOEQR, Fitness: 11796.3075445

```
SECRE   TNSAE   AVESD   ROPPI   NGISS   TILLI   NTHEN   EWSDE   TAILS   ABOUT   ONCES   ECRET
PROGR   AMSCO   NTINU   ETOLE   AKTHE   DIREC   TOROF   NATIO   NALIN   TELLI   GENCE   HASRE
CENTL   YDECL   ASSIF   IEDAD   DITIO   NALIN   FORMA   TIONA   NDTHE   PRESI   DENTS   REVIE
WGROU   PHASZ   USTRE   LEASE   DITSR   EPORT   ANDRE   COMME   NDATI   ONSWI   THALL   THISG
OINGO   NITSE   ASYTO   BECOM   EINUR   EDTOT   HEBRE   ADTHA   NDDEP   THOFT   HENSA   SACTI
VITIE   SBUTT   HROUG   HTHED   ISCLO   SURES   WEVEL   EARNE   DANEN   ORMOU   SAMOU   NTABO
UTTHE   AGENC   YSCAP   ABILI   TIESH   OWITI   SFAIL   INGTO   PROTE   CTUSA   NDWHA   TWENE
EDTOD   OTORE   GAINS   ECURI   TYINT   HEINF   ORMAT   IONAG   EFIRS   TANDF   OREMO   STTHE
SURVE   ILLAN   CESTA   TEISR   OBUST   ITISR   OBUST   POLIT   ICALL   YLEGA   LLYAN   DTECH
NICAL   LYICA   NNAME   THREE   DIFFE   RENTN   SAPRO   GRAMS   TOCOL   LECTG   MAILU   SERDA
TATHE   SEPRO   GRAMS   AREBA   SEDON   THREE   DIFFE   RENTT   ECHNI   CALEA   VESDR   OPPIN
GCAPA   BILIT   IESTH   EYREL   YONTH   REEDI   FFERE   NTLEG   ALAUT   HORIT   IESTH   EYINV
OLVEC   OLLAB   ORATI   ONSWI   THTHR   EEDIF   FEREN   TCOMP   ANIES   ANDTH   ISISZ   USTGM
AILTH   ESAME   ISTRU   EFORC   ELLPH   ONECA   LLREC   ORDSI   NTERN   ETCHA   TSCEL   LPHON
ELOCA   TIOND   ATASE   CONDT   HENSA   CONTI   NUEST   OLIEA   BOUTI   TSCAP   ABILI   TIESI
THIDE   SBEHI   NDTOR   TURED   INTER   PRETA   TIONS   OFWOR   DSLIK   ECOLL   ECTIN   CIDEN
TALLY   TARGE   TANDD   IRECT   EDITC   LOAKS   PROGR   AMSIN   MULTI   PLECO   DENAM   ESTOO
BSCUR   ETHEI   RFULL   EXTEN   TANDC   APABI   LITIE   SOFFI   CIALS   TESTI   FYTHA   TAPAR
TICUL   ARSUR   VEILL   ANCEA   CTIVI   TYISN   OTDON   EUNDE   RONEP   ARTIC   ULARP   ROGRA
MORAU   THORI   TYCON   VENIE   NTLYO   MITTI   NGTHA   TITIS   DONEU   NDERS   OMEOT   HERPR
OGRAM   ORAUT   HORIT   YTHIR   DUSGO   VERNM   ENTSU   RVEIL   LANCE   ISNOT   ZUSTA   BOUTT
HENSA   THESN   OWDEN   DOCUM   ENTSH   AVEGI   VENUS   EXTRA   ORDIN   ARYDE   TAILS   ABOUT
THENS   ASACT   IVITI   ESBUT   WENOW   KNOWT   HATTH   ECIAN   ROFBI   DEAAN   DLOCA   LPOLI
CEALL   ENGAG   EINUB   IQUIT   OUSSU   RVEIL   LANCE   USING   THESA   MESOR   TSOFE   AVESD
ROPPI   NGTOO   LSAND   THATT   HEYRE   GULAR   LYSHA   REINF   ORMAT   IONWI   THEAC   HOTHE
R
```

First result This key seemed to have one or two inversions left in it but the text was clearly about the NSA and security. So I choose a different training text, The Shadow Factory by James Bamford. After training on a text with similar words in it the average fitness of the population rose much more quickly and after convergence gave me the correct key. Below is the key and the messsage with manually added spaces.

Key BDFZKHMAWTVPGULNYJCXSIOEQR, Fitness: 12796.4907841

Message: SECRET NSA EAVESDROPPING IS STILL IN THE NEWS DETAILS ABOUT ONCE SECRET PROGRAMS CONTINUE TO LEAK THE DIRECTOR OF NATIONAL INTELLIGENCE HAS RECENTLY DECLASSIFIED ADDITIONAL INFORMATION AND THE PRESIDENTS REVIEW GROUP HAS JUST RELEASED ITS REPORT AND RECOMMENDATIONS WITH ALL THIS GOING ON ITS EASY TO BECOME INURED TO THE BREADTH AND DEPTH OF THE NSAS ACTIVITIES BUT THROUGH THE DISCLOSURES WEVE LEARNED AN ENORMOUS AMOUNT ABOUT THE AGENCYS CAPABILITIES HOW IT IS FAILING TO PROTECT US AND WHAT WE NEED TO DO TO REGAIN SECURITY IN THE INFORMATION AGE FIRST AND FOREMOST THE SURVEILLANCE STATE IS ROBUST IT IS ROBUST POLITICALLY LEGALLY AND TECHNICALLY I CAN NAME THREE DIFFERENT NSA PROGRAMS TO COLLECT GMAIL USER DATA THESE PROGRAMS ARE BASED ON THREE DIFFERENT TECHNICAL EAVESDROPPING CAPABILITIES THEY RELY ON THREE DIFFERENT LEGAL AUTHORITIES THEY INVOLVE COLLABORATIONS WITH THREE DIFFERENT COMPANIES AND THIS IS JUST GMAIL THE SAME IS TRUE FOR CELLPHONE CALL RECORDS INTERNET CHATS CELLPHONE LOCATION DATA SECOND THE NSA CONTINUES TO LIE ABOUT ITS CAPABILITIES IT HIDES BEHIND TORTURED INTERPRETATIONS OF WORDS LIKE COLLECT INCIDENTALLY TARGET AND DIRECTED IT CLOAKS PROGRAMS IN MULTIPLE CODE NAMES TO OBSCURE THEIR FULL EXTENT AND CAPABILITIES OFFICIALS TESTIFY THAT A PARTICULAR SURVEILLANCE ACTIVITY IS NOT DONE UNDER ONE PARTICULAR PROGRAM OR AUTHORITY CONVENIENTLY OMITTING THAT IT IS DONE UNDER SOME OTHER PROGRAM OR AUTHORITY THIRD US GOVERNMENT SURVEILLANCE IS NOT JUST ABOUT THE NSA THE SNOWDEN DOCUMENTS HAVE GIVEN US EXTRAORDINARY DETAILS ABOUT THE NSAS ACTIVITIES BUT WE NOW KNOW THAT THE CIA NRO FBI DEA AND LOCAL POLICE ALL ENGAGE IN UBIQUITOUS SURVEILLANCE USING THE SAME SORTS OF EAVESDROPPING TOOLS AND THAT THEY REGULARLY SHARE INFORMATION WITH EACHOTHER

4.

Vigenere Cipher The first step to solving the vigenier cipher is to find the size of the key using index of coincidence. My program (attached as solv-2.py) when passed the k option and two keysizes seperated by a colon will try seperating the text into a matrix based on that key size and calculate the IC of each column. Each column is assumed to be the letters translated by the same letter of the key if the key length is equal to the width of the matrix. The results are sorted by the IC and returned to the user to show the most probable key length first. The result of running this program is shown below.

```
$ python solv-2.py -k 3:29 hw1-2.crypt
(12, 1.707807386629266)
(20, 1.5194805194805194)
(10, 1.3886743886743886)
(24, 1.3228492136910268)
(6, 1.2981110142400465)
(21, 1.2264150943396226)
(3, 1.2167552997741677)
(4, 1.2087369815339064)
(23, 1.193877551020408)
(17, 1.187878787878788)
(8, 1.133145657387134)
.
.
.
```

Key length These results hint that the keylength is most likely 12, 20 or 10 since they have the highest index of coincidence. ga-2.py was used to test a keylength of 12 and see if it could find any keys that yielded english looking text using the exact same technique from the first decryption problem. Again the GA was trained on The Shadow Factory because I assumed that the text had similar content to the first one.

```
$ python ga-2.py -l 12 hw1-2.crypt
Geneation 1 best individual: (YTLHJKJOEUUH) 3048.749782
Geneation 2 best individual: (AFZVIFBEWJUF) 3927.156567
Geneation 3 best individual: (ARZVIFJFKUUH) 4671.765319
Geneation 4 best individual: (AFZVIFBEKUUH) 6447.459217
Geneation 5 best individual: (RXFFIFBEKUUH) 6554.993247
Geneation 6 best individual: (AXFVIFBEKUUH) 7835.351044
Geneation 7 best individual: (AFFGVFBEKUUH) 8194.710882
Geneation 8 best individual: (AXFGVFBEKUUH) 9500.430740
Geneation 9 best individual: (AXFGVFBEKUUH) 9500.430740
```

Solution The GA converged after 8 generations giving us the key YTLHJKJOEUUH. It looks like the GA got a little lucky successfully guessing the last three letters in the first generation. After pulling the data out of the gen8 file I fixed the spacing to by hand. The final result is shown below:

Message: PRESIDENT BARACK OBAMA IS EXPECTED TO ANNOUNCE CHANGES ON FRIDAY TO SWEEPING US SURVEILLANCE EFFORTS EXPOSED BY INTELLIGENCE LEAKER EDWARD SNOWDEN WHOSE BLOCKBUSTER DISCLOSURES HAVE RAISED QUESTIONS ABOUT GOVERNMENT OVERREACH IN FIGHTING TERROR THE SCOPE OF PHONE AND EMAIL SNOOPING BY THE NATIONAL SECURITY AGENCY THAT CAME TO LIGHT LAST YEAR TRIGGERED OUTRAGE FROM CIVIL LIBERTARIANS AND PROMPTED KEY MEMBERS OF CONGRESS FROM BOTH PARTIES TO WEIGH CHANGES IN NATIONAL SECURITY LAW OBAMA IS EXPECTED TO ACT ON RECOMMENDATIONS FROM AN INDEPENDENT PANEL THAT HE CALLED FOR AT THE HEIGHT OF THE FALLOUT FROM THE LEAKS AROUND THE AGENCYS SURVEILLANCE ACTIVITIES AND A SECRET COURT

THAT WORKS WITH IT THE PRESIDENT IAL REVIEW GROUP ON INTELLIGE NCE CONCLUDED IN DECEMBER THAT DATA COLLECTION SHOULD REMAIN BUT THAT THE GOVERNMENT MUST DO A BETTER JOB OF PROTECTING CIVIL LIBERTIES IN THE CONTEXT OF NATIONAL SECURITY CHANGES IMPOSED BY THE PRESIDENT WILL PERMANENTLY PLACE HIS SIGNATURE ON THE INTELLIGENCE INITIATIVE AND HELP DEFINE HIS LEGACY AS A CHIEF EXECUTIVE WHO PROMISED A MORE OPEN AND TRANSPARENT GOVERNMENT WHEN HE ENTERED THE WHITE HOUSE FIVE YEARS AGO NSA DOMESTIC AND INTERNATIONAL PHONE AND EMAIL SURVEILLANCE IS CONS IDEREDSOME OF THE MOST WIDESPREAD INTELLIGENCE GATHERING PERFORMED BY THE US GOVERNMENT

5.

Bible 1 The bible has ~ 31000 verses. So we have that many starting places, and we know the length of the cipher text. It should be a trivial matter to extract the text starting at each verse and keep reading as long as there is more ciphertext. I think it would be $\mathrm{O}\left(n\right)$ where n is the length of the ciphertext. Not too difficult.

Bible 2 Given that we do not know the function to choose the next verse (whereas in the last one we did, it was simply the next verse) it should be a lot harder to figure out the key. Assuming of course that the ciphertext is long enough to span more than one verse we will have to calculate the next verse each time we get to the end of a verse but we still have ciphertext without a key. If we were to brute force it we could simply try all verses, then try all verses as the second verse in the key, then all verses as the third and so on. However this is going to get difficult very fast since we are multiplying by ~ 31000 each time.

6.

Entropy It does not. Entropy is a measure of the information content of the message. As long as a message is recoverable given a ciphertext and a key then the message still contains the content, and therefore it still contains the same entropy.