1 Playfair

1.1 De opdracht

De tweede ciphertekst was versleuteld met Playfair. We moesten dus de key zien te vinden waarmee een matrix was opgesteld om digrammen te versleutelen. We hadden op dit punt al een Nederlandse en een Franse tekst ontcijferd, dus we gingen ervanuit dat dit Engels was. Het oplossen van Playfair bleek moeilijker als ADFGVX en Vigenere, na 4 gefaalde pogingen is het ons uiteindelijk gelukt.

1.2 Poging 1: frequentie analyse

Onze eerste poging bestond uit een frequentieanalyse van de digrammen. We vergeleken die waarden met gekende waarden voor digrammen in het Engels die we zelf hadden berekend aan de hand van enkele Engelse plainteksten. Na veel puzzelwerk raakten we echter niet erg ver. Er zijn 600 digrammen in Playfair (hoewel ze niet allemaal in de ciphertekst voorkwamen) en door dicht bij elkaar gelegen frequenties was het bijna onmogelijk ze juist te kiezen. We bekeken ook de frequenties van opeenvolgende digrammen samen en we hielden rekening met frequente digrammen waarvan ook het omgekeerde frequent was.

1.3 Poging 2: slimme frequentie analyse

Aangezien we door gewoon naar de frequenties te kijken toch 2 digrammen vrij zeker wisten ("OS" $\leftrightarrow$ "th", "OB" $\leftrightarrow$ "he"), probeerden we om gebruik te maken van de structuur van het Playfair vierkant[1]. Dit werkte echter ook niet, omdat we geen van de andere digrammen echt zeker wisten, en diegenen die we dachten juist te gokken bleken niet in het vierkant te passen.

Dus probeerden we de ciphertekst te bruteforcen met keys waarbij deze reeds gevonden digrammen klopten, maar vonden ook zo het antwoord niet.

1.4 Poging 3: Hill Climb Algoritme

Daarna gooiden we het over een andere boeg en gingen we op zoek naar een algoritme om de ciphertekst volledig geautomatiseerd te kraken. Eerst maakten we een implementatie van een hill climb algoritme. We starten met een bepaalde startset van mogelijke keys (in ons geval 100 random gekozen keys), de besten op basis van Index of Coincedence[2][3] hieruit te kiezen en deze te combineren, en deze combinaties dan terug als nieuwe startset te

[1] `http://www.umich.edu/~umich/fm-34-40-2/ch7.pdf`

[2] `http://practicalcryptography.com/cryptanalysis/text-characterisation/index-coincidence/`

[3] `http://en.wikipedia.org/wiki/Index_of_coincidence`

beschouwen. Op een gegeven moment zitten we in een maximum (lokaal) en dit wordt dan teruggegeven. Doordat we voor de combinatie-stap kozen voor een groot aantal keer 2 letters in een key te wisselen en er dus eigenlijk geen combinatie plaatsvond duurde dit algoritme extreem lang om iets te vinden of belande het direct in een lokaal en zeer slecht maximum.

1.5 Poging 4: Churn Algoritme

Daarna stootten we op het zogenaamde Churn algoritme [4] [5]. Dit algoritme is gelijkaardig aan simulated annealing, alleen heel wat simpeler om te implementeren.
Elke plaintekst krijgt een score toegewezen als volgt: voor elke mogelijke digram in het Engels werd de frequentie geanalyseerd. Hiervan werden telkens de log genomen om de invloed van erg grote waarden te beperken, en werden deze waarden herschaald naar 0-9. Nu wordt voor elk digram (ook twee opeenvolgende letters die volgens Playfair in verschillende digrams zitten!) in de plaintekst de frequentie van 0-9 bij de score van de plaintekst geteld. Hoe hoger de score, hoe waarschijnlijker dat de tekst Engels is dus. Merk hierbij op dat de plainteksten van de gegeven ciphertekst komen, dus erg hoog scorende maar niet-Engelse plainteksten als "eeeeeeeeee" zullen niet voorkomen. In het algoritme starten we met een zogenaamde parent key (bv. gewoon het alfabet) waarmee we de ciphertekst ontcijferen en een score toekennen. Daarna voeren we een kleine verandering door aan de parent key en noemen we het resultaat de child key. Deze verandering kan een horizontale of verticale spiegeling, een combinatie van de twee, een verwisseling van twee willekeurige rijen of kolommen of een verwisseling van twee letters zijn. Hierbij komt de verwisseling van twee letters veel meer voor dan de andere opties, dit omdat er veel meer mogelijke verwisselingen van twee letters zijn. De plaintekst voor de child key wordt ook geëvalueerd. Indien de child key beter scoorde, vervangt deze de parent key. Indien de parent key beter scoorde, wordt een willekeurig getal uit een array van 100 getallen gekozen. Indien het verschil tussen parent en child key scores minder was dan dit gekozen getal, vervangt de child key toch de parent key. Hierdoor kan het algoritme uit lokale maxima raken. Het algoritme blijft oneindig lopen en print de uitkomst van een iteratie enkel indien een nieuwe topscore bereikt is. Merk op dat de 100 getallen in de genoemde array zodanig gekozen zijn dat de kans dat child parent vervangt, gelijkaardig is aan die bij simulated annealing. Bij sommige runs van het algoritme vonden we al na een tweeduizendtal iteraties een tekst die heel erg op Engels leek. Het antwoord was niet helemaal correct gezien bij de logaritmen van de frequenties van digrammen geen rekening gehouden werd met de meer voorkomende X bij

[4] http://www.cryptoden.com/index.php/algorithms/churn-algorithm/20-churn-algorithm

[5] http://s13.zetaboards.com/Crypto/topic/6781204/1/

Playfair. Het was wel dicht genoeg bij Engels dat we de laatste aanpassingen handmatig konden doorvoeren. We vonden als key "A brief history of time", met als plaintekst het begin van "A Brief History of Time: From the Big Bang to Black Holes" door Stephen Hawking. [6]. Merk op dat we de twee lijsten met waarden op `http://www.cryptoden.com/` vonden, maar de code verder helemaal zelf geschreven is en enkel gebaseerd is op de beschrijving van het algoritme.
Volgens de beschrijving van het algoritme waren de 100 waarden in de array afgesteld op een ciphertekst van 100-120 karakters en moest deze nog geschaald worden voor langere teksten. Wij kregen echter betere resultaten zonder de waarden te schalen.
Het algoritme in playfair.py is een licht aangepaste versie van wat we eerst gebruikten. Het geeft vaker snel een antwoord door rollbacks uit te voeren als een tijdje geen vooruitgang geboekt wordt, of zelfs helemaal opnieuw te beginnen. Ook is het scoren nu aangepast aan de X'en in Playfair waardoor exact de correcte plaintekst wordt gevonden. De gevonden key matrix is niet altijd de matrix gegenereerd met "abriefhistoryoftime", maar wel altijd een correcte matrix. Gezien niet alle digrammen voorkomen, zijn meerdere correcte matrices mogelijk.
In de eerste twee lijntjes van het Churn algoritme wordt een vaste seed ingesteld. Deze levert al na 2406 iteraties het antwoord op. Om met een andere seed te proberen, moeten de eerste twee regels code verwijderd worden. Hierdoor kan het algoritme wel langer duren. De variërende runtime is een gevolg van de niet-deterministische aard van het algoritme. Gezien de code niet concurrent is, is het aangewezen om het programma meerdere keren tegelijk te draaien om sneller tot een resultaat te komen.

1.6 De encrypted text

AWELXLKNOWNSCIENTISTSOMESAYITWASBERTRANDRUSXSEL
XLONCEGAVEAPUBLICLECTUREONASTRONOMYHEDESCRIBED
HOWTHEXEARTHORBITSAROUNDTHESUNANDHOWTHESUNINT
URNORBITSAROUNDTHECENTEROFAVASTCOLXLECTIONOFST
ARSCALXLEDOURGALAXYATXTHEXENDOFTHELECTUREALITX
TLEOLDLADYATXTHEBACKOFTHEROXOMGOTUPANDSAIDWH
ATYOUHAVETOLDUSISRUBXBISHTHEWORLDISREALXLYAFLAT
PLATESUPXPORTEDONTHEBACKOFAGIANTXTORTOISETHESCI
ENTISTGAVEASUPERIORSMILEBEFOREREPLYINGWHATISTHET
ORTOISESTANDINGONYOUREVERYCLEVERYOUNGMANVERYCL
EVERSAIDTHEOLDLADYBUTITSTURTLESALXLTHEWAYDOWNM
OSTPEOPLEWOULDFINDTHEPICTUREOFOURUNIVERSEASANIN
FINITETOWEROFTORTOISESRATHERXRIDICULOUSBUTWHYDO

[6] `http://www.fisica.net/relatividade/stephen_hawking_a_brief_history_of_time.pdf#page=3`

WETHINKWEKNOWBETXTERWHATDOWEKNOWABOUTXTHEU
NIVERSEANDHOWDOWEKNOWITWHEREDIDTHEUNIVERSECOM
EFROMANDWHEREISITGOINGDIDTHEUNIVERSEHAVEABEGINX
NINGANDIFSOWHATHAPXPENEDBEFORETHENWHATISTHENAT
UREOFTIMEWILXLITEVERCOMETOANENDCANWEGOBACKINTI
MERECENTBREAKTHROUGHSINPHYSICSMADEPOSXSIBLEINPA
RTBYFANTASTICNEWTECHNOLOGIESXSUGXGESTANSWERSTO
SOMEOFTHESELONGSTANDINGQUESTIONSXSOMEDAYTHESEA
NSWERSMAYSEXEMASOBVIOUSTOUSASTHEXEARTHORBITING
THESUNORPERHAPSASRIDICULOUSASATOWEROFTORTOISESO
NLYTIMEWHATEVERTHATMAYBEWILXLTELXLASLONGAGOAS
THREXEHUNDREDANDFOURTYBCTHEGREXEKPHILOSOPHERA
RISTOTLEINHISBOXOKONTHEHEAVENSWASABLETOPUTFORW
ARDTWOGOXODARGUMENTSFORBELIEVINGTHATXTHEXEART
HWASAROUNDSPHERERATHERTHANAHATPLATEFIRSTHEREAL
IZEDTHATECLIPSESOFTHEMOXONWERECAUSEDBYTHEXEART
HCOMINGBETWEXENTHESUNANDTHEMOXONTHEXEARTHSXSH
ADOWONTHEMOXONWASALWAYSROUNDWHICHWOULDBETRU
EONLYIFTHEXEARTHWASXSPHERICALIFTHEXEARTHXHADBEX
ENAFLATDISKTHESHADOWXWOULDHAVEBEXENELONGATEDA
NDELXLIPTICALUNLESXSTHEXECLIPSEALWAYSOCXCURXRED
ATATIMEWHENTHESUNWASDIRECTLYUNDERTHECENTEROFT
HEDISKSECONDTHEGREXEKSKNEWFROMTHEIRTRAVELSTHAT
XTHENORTHSTARAPXPEAREDLOWERINTHESKYWHENVIEWED
INTHESOUTHTHANITDIDINMORENORTHERLYREGIONSXSINCET
HENORTHSTARLIESOVERTHENORTHPOLEITAPXPEARSTOBEDI
RECTLYABOVEANOBSERVERATXTHENORTHPOLEBUTXTOSOM
EONELOXOKINGFROMTHEXEQUATORITAPXPEARSTOLIEIUSTA
TXTHEHORIZONFROMTHEDIFXFERENCEINTHEAPXPARENTPO
SITIONOFTHENORTHSTARINEGYPTANDGREXECEARISTOTLEX
EVENQUOTEDANESTIMATETHATXTHEDISTANCEAROUNDTHEX
EARTHWASFOURHUNDREDTHOUSANDSTADIAITISNOTKNOWNE
XACTLYWHATLENGTHASTADIUMWASBUTITMAYHAVEBEXENA
BOUTXTWOHUNDREDYARDSWHICHWOULDMAKEARISTOTLESE
STIMATEABOUTXTWICETHECURXRENTLYACXCEPTEDFIGURE
THEGREXEKSEVENHADATHIRDARGUMENTXTHATXTHEXEART
HMUSTBEROUNDFORWHYELSEDOESONEFIRSTSEXETHESAILSO
FASHIPCOMINGOVERTHEHORIZONANDONLYLATERSEXETHEH
ULXLARISTOTLETHOUGHTXTHEXEARTHWASXSTATIONARYAN
DTHATXTHESUNTHEMOXONTHEPLANETSANDTHESTARSMOVE
DINCIRCULARORBITSABOUTXTHEXEARTHXHEBELIEVEDTHISB
ECAUSEHEFELTFORMYSTICALREASONSTHATXTHEXEARTHWA
STHECENTEROFTHEUNIVERSEANDTHATCIRCULARMOTIONWA
STHEMOSTPERFECTXTHISIDEAWASELABORATEDBYPTOLEMY

INTHESECONDCENTURYADINTOACOMPLETECOSMOLOGICALM
ODELTHEXEARTHSTOXODATXTHECENTERSURXROUNDEDBYE
IGHTSPHERESTHATCARXRIEDTHEMOXONTHESUNTHESTARSA
NDTHEFIVEPLANETSKNOWNATXTHETIMEMERCURYVENUSMA
RSIUPITERANDSATURNX