Fall 2015

Modern Crypto-Analysis of Polyalphabetic Ciphers Using a Genetic Algorithm

Benjamin E. Heald '19, Gettysburg College

Follow this and additional works at: http://cupola.gettysburg.edu/student_scholarship

Part of the Mathematics Commons

Share feedback about the accessibility of this item.

http://cupola.gettysburg.edu/student_scholarship/405

This is the author's version of the work. This publication appears in Gettysburg College's institutional repository by permission of the copyright owner for personal use, not for redistribution. Cupola permanent link: http://cupola.gettysburg.edu/student_scholarship/405

This open access poster is brought to you by The Cupola: Scholarship at Gettysburg College. It has been accepted for inclusion by an authorized administrator of The Cupola. For more information, please contact cupola@gettysburg.edu.
Modern Crypto-Analysis of Polyalphabetic Ciphers Using a Genetic Algorithm

Abstract
This project involved implementing a genetic algorithm to help automate the crypto-analysis of the Vigenere cipher.

Keywords
Cryptography, Genetic, Algorithm, Vigenere

Disciplines
Mathematics

Comments
This poster was made for Professor Darren Glass's First Year Seminar, FYS 146-2: Cryptography: The Science of Secrecy, Fall 2015. It was presented as part of the first CAFE Symposium, 2016.
Modern Crypto-Analysis of Polyalphabetic Ciphers Using a Genetic Algorithm
Benjamin Heald

The Vigenere Cipher

- Later misattributed to Blaise de Vigenère
- Was considered for hundreds of years to be “unbreakable”
- Broken in 1854 by Charles Babbage.
- Is a combination of many so-called mono-alphabetic ciphers.

Step One: Finding the Key Length
- Kasiki Test to find the most likely candidates, i.e., 2, 4, 8.
- These candidates are then analyzed by the Friedman Method to determine which most resembles English text.
- The standard “Index of coincidence” for English text is 1.73
- The candidate with the “IC” closest to this ideal is the most likely key length.
- This combination of the two algorithms makes it much easier for a basic computer program to find the most likely key length.

Step Two: Finding the Key Text
- With a known key length, a variety of attacks may now be used.
- Without human interaction however, it is very difficult for the computer to know if it has arrived at a solution.

Method One: Brute Force Attack
Method Two: Common-Word Dictionary Attack
Method Three: Markov Chain
Method Four: A Genetic Algorithm

Key Length and Speed

- This algorithm finds almost 70% of the correct letters in keywords of length greater than four.
- With further optimization however, this number reaches 98%.
- This experiment was carried out on an older model PC, with newer hardware, much faster speeds could be achieved.

Summary

- Through combining many different algorithms, the key length can be found very reliably.
- The Genetic Algorithm is best used when combined with the Friedman Method.
- The Genetic Algorithm, once properly implemented, is the best and most accurate way to automate the decipherment process.
- Without human interaction, the entire process becomes much harder.
- Though not perfect, this program is relatively accurate for all key lengths.
- With greater key lengths, the more genetic “generations” are needed to arrive at the solution.
- Automating the crypto-analysis makes finding the key much faster than when done by hand.
- With newer hardware, the analysis could be done even faster.
- When incorporated into a web interface, the program is very accessible.

Acknowledgments

- Professor Glass for his First-Year seminar on the science and history behind Cryptography.