Cryptanalysis using Genetic Algorithms

Karel P. Bergmann
Dept. of Computer Science
University of Calgary
2500 University Dr. NW
Calgary, AB, Canada
kpbergma@ucalgary.ca

Renate Scheidler
Dept. of Mathematics
University of Calgary
2500 University Dr. NW
Calgary, AB, Canada
rscheidl@math.ucalgary.ca

Christian Jacob
Dept. of Computer Science
University of Calgary
2500 University Dr. NW
Calgary, AB, Canada
cjacob@ucalgary.ca

ABSTRACT

Genetic algorithms were used for the cryptanalysis of a number of classical cryptosystems. The results of applying a number of GA implementations to various ciphers are presented. The selection of appropriate mutation operators and fitness functions is also discussed.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Heuristic methods*

General Terms

Algorithms, Security

1. OVERVIEW

This work applied specialized genetic algorithms [6] to attack a number of different *cryptosystems* in order to ascertain their security against such attacks. It was found that three different classical substitution ciphers (Vigenère, Mixed Vigenère and autokey) were highly susceptible to cryptanalysis by genetic algorithms. Although similar results were attainable against the columnar transposition cipher using the same genetic algorithm implementation, it was found that far better results could be obtained by using a genetic algorithm tailored to the task of evolving permutations. Finally, a genetic algorithm was applied to the cryptanalysis of two modern encryption standards (DES and AES) with no success.

2. PREVIOUS WORK

The great majority of the literature on genetic algorithm cryptanalysis attempts to apply the technique to a single cryptosystem. The research presented here included a comparative analysis of the effectiveness of different mutation operators, but also, as is presented in this paper, compared the vulnerability of a variety of ciphers to similar genetic algorithm attacks. Previous work in this area is nicely summarized in [2].

3. EXPERIMENTS

The GA described below was applied to the cryptanalysis of the Vigenère, Mixed Vigenère and Autokey ciphers [7]. Encryption and decryption of these cryptosystems operate by bijectively substituting letters in the message with other letters in order to produce a ciphertext. A key word of variable length dictates the substitutions. In the context of this research we considered cryptanalysis to be the act of re-covering a secret message given only the ciphertext, but without knowledge of the secret key which was used to produce this ciphertext.

3.1 Mutation Operators

Four mutation operators were used to cryptanalyze the substitution ciphers. Alternate mutation operators for the manipulation of permutations were developed to support a permutation-based, as opposed to string-based GA.

- *resize* (RS) - Chooses a position in a key and either removes the letter or inserts a random letter
- *character flip* (CF) - Chooses a position in a key and replaces the letter with a random letter
- *mutate by one* (MO) - Chooses a position in a key and replaces the letter with a neighbouring letter in the alphabet
- *crossover* (CO) - Replaces a random interval within one key with the corresponding letters of another key

3.2 Fitness Function

The fitness functions used throughout this research were predicated on the assumption that the ciphertext which was being cryptanalyzed decrypted to a message which was written in English. Two functions were used to form a single aggregate fitness function in this research. The first fitness function was based on the ϕ-statistic. The ϕ-statistic is a measure of the redundancy found in a piece of text. English, like other languages has its own characteristic redundancy [5]. This first fitness function took an individual for the population (key) and decrypted the ciphertext being cryptanalyzed using that key. The function then calculated the ϕ-statistic of this decryption and compared it to the expected ϕ-statistic of a piece of English text of the same length and a piece of random text of the same length. The second fitness function depended on the use of *cribs*, or pieces of known text appearing in a message. This fitness function assigned a fitness between 0 and 1 to an individual (key) by decrypting the ciphertext with the key and then checking for the presence of the supplied crib at a predetermined location in the decrypted ciphertext, using Hamming Distance.

GECCO'08, July 12–16, 2008, Atlanta, Georgia, USA.
ACM 978-1-60558-130-9/08/07.

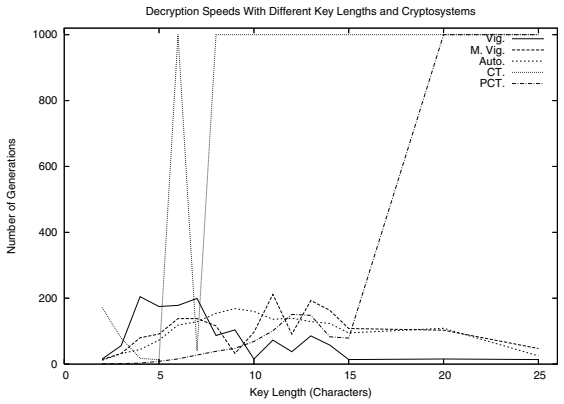

Figure 1: Decryption Speeds With Different Key Lengths and GA Implementations

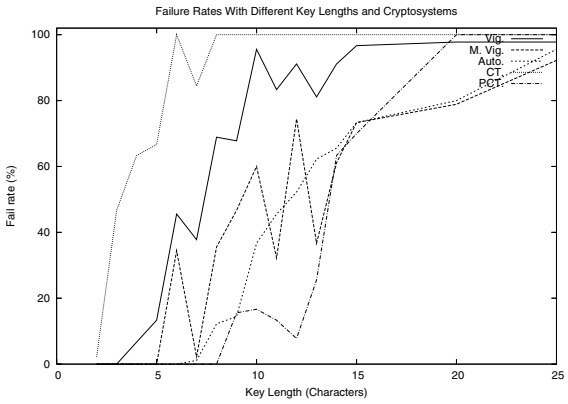

Figure 2: Failure Rates With Different Key Lengths and GA Implementations

3.3 Results

The GA was tested on all three of the polyalphabetic substitution ciphers described previously. Nine different combinations of messages and ciphertexts were used, and each combination was cryptanalyzed 10 times for each cryptosystem using key lengths ranging from 2 to 25. The following list contains all of the parameter settings used by the GA.

- Population size: 20
- Number of individuals tenured per generation: 5
- Number of random immigrants per generation: 5
- Maximum number of generations: 1000
- Maximum key length: 1.5 × actual key length
- Ciphertext length: 500 characters

As can be seen in Figures 1 and 2, all of the substitution ciphers behaved similarly during these tests. If the GA succeeded in cryptanalyzing the ciphertext, then it did so within 200 generations. As the key length increased, the number of generations required for cryptanalysis did not increase, but the likelihood of failure did. Cryptanalysis with a key lengths below 7 was routine, while cryptanalysis with key lengths above 15 became nearly impossible.

We also applied our GA to the columnar transposition cipher [7], a system that permutes the letters of a message to produce a ciphertext. Here, the results (CT) turned out to be quite poor, but greatly improved when switching to a permutation-based GA (PCT). The permutation-based GA made it possible to correctly decipher a columnar transposition ciphertext with a key of up to 12 characters routinely. As with the substitution ciphers, it can be seen in Figures 1 and 2 that longer keys didn't necessarily result in a larger number of required generations, but only in a lower likelihood of success. In addition to the previously mentioned classical cryptosystems, we attempted to apply genetic algorithm cryptanalysis to two modern ciphers, namely the Data Encryption Standard (DES) [3] and the Advanced Encryption Standard (AES) [4]. Both ciphers operate using blocks of bits as input, output and keys. The general function of both cryptosystems consists of the repeated application of alternating permutations and substitutions during a specified number of rounds. Not surprisingly, even low round variants of DES proved to be resistant to GA cryptanalysis.

4. CONCLUSION

A number of genetic algorithm implementations have been introduced, all of them tailored to the cryptanalysis of a certain type of cryptosystem. These GA's were applied to the cryptanalysis of a number of polyalphabetic substitution ciphers, a transposition cipher and two modern product ciphers. The approach was found to be effective against the polyalphabetic substitution ciphers. All three of these ciphers behaved similarly as the length of the key used was increased. It was found that in general the genetic algorithm could identify the correct key within 200 20-individual generations. The likelihood of success was very good until a key length of 7 was reached, after which point the success rate decreased rapidly. A permutation-based GA was developed to better attack transposition ciphers with great success. A bit-based GA was developed for application to DES and AES without the same measure of success.

5. REFERENCES

[1] K. P. Bergmann *Cryptanalysis Using Nature-Inspired Optimization Algorithms* University of Calgary, 2006, Technical Report 2008-894-07

[2] B. Delman, *Genetic Algorithms in Cryptography.* Rochester Institute of Technology, 2004

[3] National Institute of Standards and Technology, *FIPS Publication 46: Announcing the Data Encryption Standard.* FIPS, Jan. 1977

[4] National Institute of Standards and Technology, *FIPS Publication 197: Announcing the Advanced Encryption Standard (AES).* FIPS, Nov. 2001

[5] W. Friedman, *The Index of Coincidence and Its Applications in Cryptography.* Riverbank Labs, 1920

[6] J. H. Holland, *Adaptation in Natural and Artificial Systems.* The University of Michigan Press, Ann Arbor, MI, 1975

[7] A. J. Menezes and P. C. van Oorschot and S. A. Vanstone, *Handbook of Applied Cryptography.* CRC Press, Boca Raton, FL, 1997