# COMP 550- Project Proposal

Victor Redko, Marie Payne 260686859

October 23, 2018

## 1 Proposal

In classical cryptography, running key ciphers are a form of substitution cipher where typically an English phrase or word is used to provide a keystream as input to a substitution function with a plaintext to produce an output ciphertext. For simplification, the plaintext is usually also an English phrase or word and the function used is usually the *tabula recta* substitution function, defined as c = (p + r) mod 26, where c, p and r are ciphertext, plaintext and keystream letters respectively. The alphabet is 0-indexed in this system. If the keystream were truly random, this reduces to the one-time pad, a method which is proven to be unbreakable. Since the running key cipher operates under the assumption that the key, as well as the message, contain patterns typical of the English language, it can be broken using natural language processing methods.

Previously, solutions for decoding running key ciphers with English keystreams and plaintext have been proposed using Gibbs sampling (Knight and Reddy, 2012), Viterbi decoding (Griffing, 2006), and n-gram distributions without smoothing (Bauer and Tate, 2002), with Gibbs sampling outperforming Viterbi and unsmoothed n-gram solutions. Using ciphertexts of length 1000, Knight states that Gibbs performs accurately at the rate of around 90%, Viterbi around 60%, and unsmoothed n-grams around 30%. Knight also found that performance was tied to the training corpus, with the Wall Street Journal corpora outperforming the Project Gutenberg corpora by approximately 4-5%. We propose to implement these solutions and replicate their results, and hypothesize methods of obtaining better performing results using letter and word n-gram methods.

## 2 References

Bauer, C. and Tate, C. (2002). A statistical attack on the running key cipher. *Cryptologia Volume 26, Issue 4.*
Griffing, A. (2006). Solving the Running Key Cipher with the Viterbi Algorithm. *Cryptologia Volume 30, Issue 4*, pages 361-367.
Knight, K. and Reddy, S. (2012). Decoding Running Key Ciphers. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 80-84.