Plaintext Recovery for One-Time Pads Used Twice

Gregory V. Bard,
University of Wisconsin—Stout
bardg@uwstout.edu

Theodore McDonough,
University of Wisconsin—River Falls

Abstract

The one-time pad is a very simple encryption scheme taught in most first-year cryptography courses. It consists of a pad $\vec{k} = (k_1, k_2, \ldots, k_\ell)$, a sequence of independently uniformly random elements of the integers mod n, in the possession of both the sender and the receiver. The sender encodes the plaintext $\vec{p} = (p_1, p_2, \ldots, p_\ell)$ as a sequence of symbols mod n. Encryption and decryption are simply addition and subtraction mod n. More precisely,

$$c_i = p_i + k_i \bmod n.$$

The cipher is provably secure, but the classical proof makes explicit assumptions about the method of use. In particular, each pad (each $\vec{k}$) must be used only once—hence the name "one-time pad." From the National Security Agency (NSA) Venona project, it has become known that the cipher can be broken if a single pad is used twice. For example, if $\vec{c_1} = \vec{p_1} + \vec{k}$ and $\vec{c_2} = \vec{p_2} + \vec{k}$, and both $\vec{c_1}$ and $\vec{c_2}$ are intercepted, historical records indicate that it must be the case that it is computationally feasible to recover $\vec{p_1}$, $\vec{p_2}$, and $\vec{k}$. Doing so can be called the "two-time pad problem." (In fact, such cryptanalysis lead to the arrest of Julius and Ethel Rosenberg.)

The authors propose a method, based on matrix computations mod n, which efficiently solves the "two-time pad problem" for plaintexts chosen from the English language, and working with $n = 29$. The method should be effective for other modern spoken languages and similar sized n. Previous approaches to this problem assume some knowledge of $\vec{p_1}$ or $\vec{p_2}$, i.e. a "known-plaintext attack" or "known-format attack," or that the pad was used many times. For example, a single pad used for each of 10 different plaintexts was a homework problem in a course on Cryptography by Prof. Dan Boneh on *Coursera*. Here, only the probability distribution of the underlying language is used.

This is joint work with Theodore McDonough, an undergraduate doing research under the McNair Program at the University of Wisconsin—River Falls.

Keywords

ACA 2014, Cryptography, One-Time Pad, Matrices mod n, Cryptanalysis, Two-Time Pad Problem.