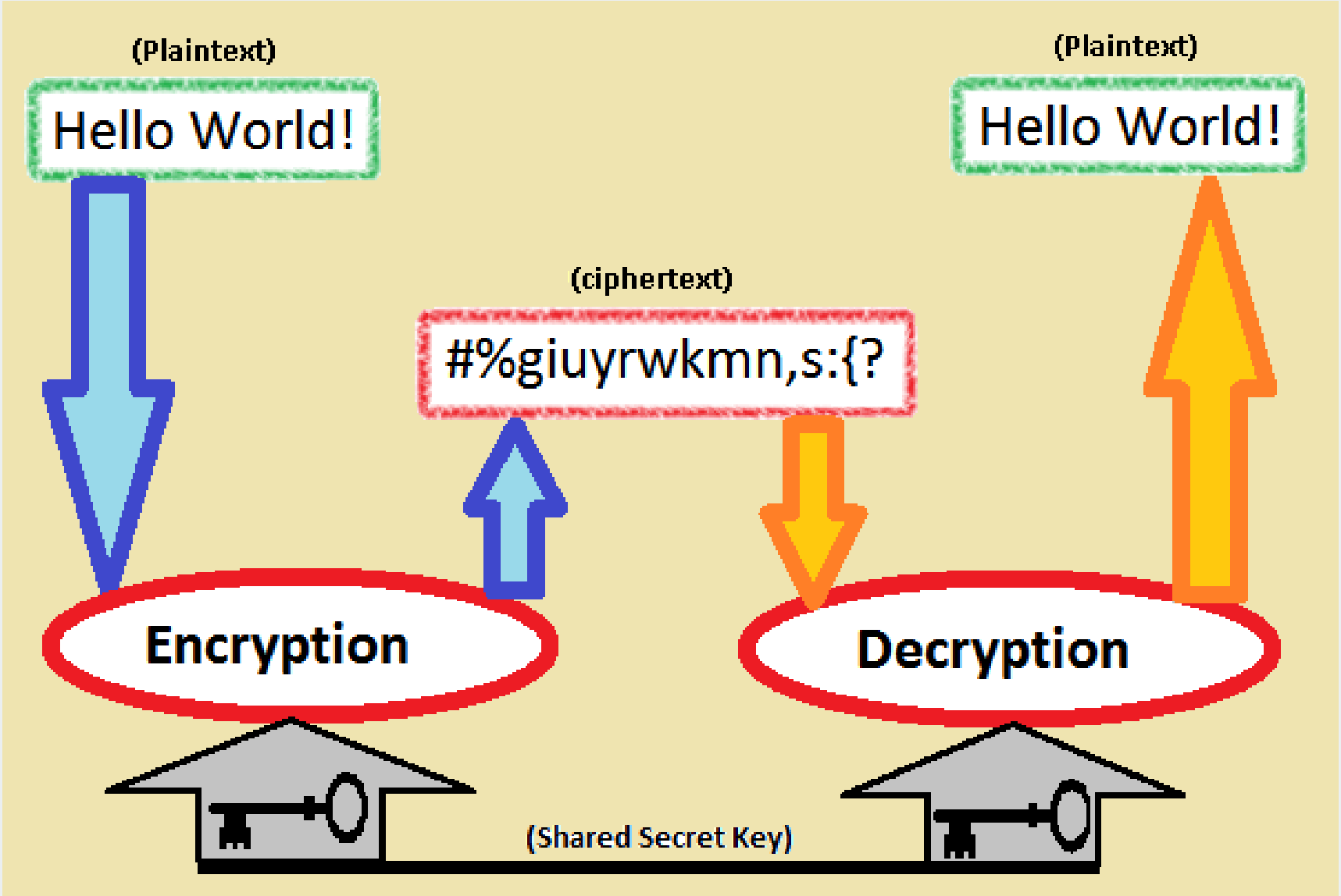
Spurious Keys and Unicity Distance

Ana Smaranda Sandu
Advisor: Professor Randy Shull, Computer Science Department

Introduction

Cryptographic System

A cryptographic system is composed of a set of plaintexts, a set of ciphertexts, a set of possible keys and a pair of encryption and decryption rules for each key.



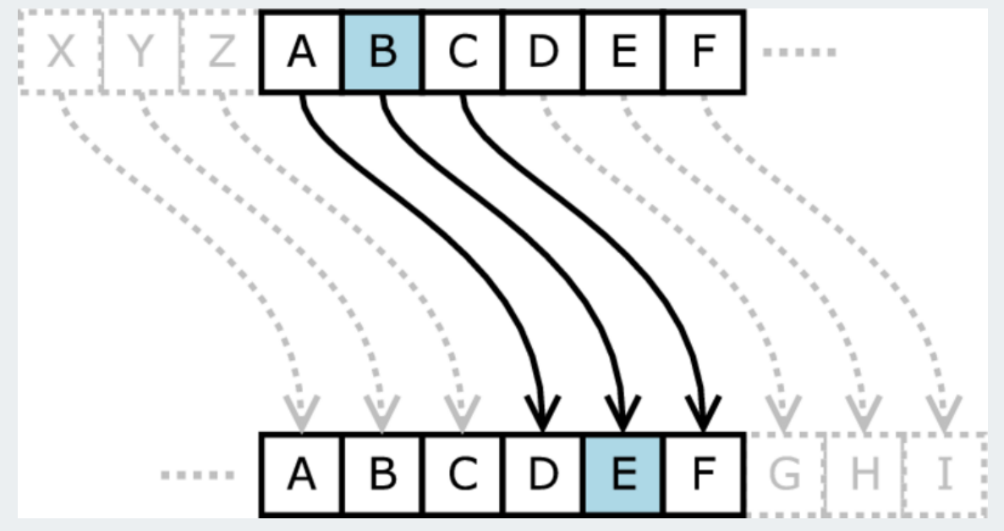
Source: http://upload.wikimedia.org/wikipedia/commons/f/f8/Crypto.png

In particular, a cryptosystem is a five-tuple $(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$ where $\mathcal{P}$ is a finite set of possible plaintexts, $\mathcal{C}$ is a finite set of possible ciphertexts and $\mathcal{K}$, the keyspace, is a finite set of possible keys. For each $K \in \mathcal{K}$, there is an encryption rule, $e_K \in \mathcal{E}$ and a corresponding decryption rule $d_K \in \mathcal{D}$ such that $d_K(e_K(x)) = x$ for every plaintext element $x \in \mathcal{P}$.

Perfect Secrecy

A cryptosystem has perfect secrecy if the a posteriori probability that a plaintext is x given an observed ciphertext y is identical to the a priori probability that the plaintext is x.

Example. Consider the shift cipher.



Source: http://51713941.weebly.com/history−of−cryptology.html

Suppose $K = c$, which is the third letter in the alphabet. Let our ciphertext be “wewillmeet". Then the corresponding positions of these letters in the alphabet is

$$22\ 4\ 22\ 8\ 11\ 11\ 12\ 4\ 4\ 19.$$

Then adding 3 to each value gives

$$25\ 7\ 25\ 11\ 14\ 14\ 15\ 7\ 7\ 22,$$

which becomes

$$ZHZLOOMHHW.$$

Fun fact: The shift cipher has perfect secrecy for shifts of one character!

Entropy

Entropy is a measure of uncertainty and of how much information can be stored in a unit, so that we can accurately represent all outcomes of an event.

Definition. *Suppose $\mathbf{X}$ is a discrete random variable which takes on values from a finite set X. Then the entropy of the random variable $\mathbf{X}$ is defined to be the quantity*

$$H(\mathbf{X}) = -\sum_{x \in X} Pr[x] \log_2 Pr[x].$$

Entropy and Redundancy

Question: How much information can a language store?
Answer: We measure this by H_L, the entropy per letter of a natural language. This is the average information per letter in a meaningful string of text.

Approximation #1: $H(P)$ - the entropy of the random variable associated with the plaintexts.
Approximation #2: $H(P^n)$ - the entropy of the random variable representing plaintexts of length n.

Definition. *Suppose L is a natural language. The entropy of L is defined to be the quantity*

$$H_L = \lim_{n\to\infty} \frac{H(P^n)}{n},$$

where P^n is the random variable that has its probability distribution that of all plaintexts of length n. We also define the redundancy of L to be given by

$$R_L = 1 - \frac{H_L}{\log_2 |\mathcal{P}|}.$$

Fun fact: $R_L = 0.75$ for the English language, so the English language is 75% redundant!

Spurious Keys

Suppose we have a cryptosystem and a plaintext x encrypted with a key k resulting in ciphertext y. Knowing only the ciphertext y, how can we determine the key? These keys that we cannot eliminate by any standard approach are called spurious keys.

Theorem. *Suppose $(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$ is a cryptosystem where $|\mathcal{C}| = |\mathcal{P}|$ and keys are chosen with the same probability. Then given a ciphertext of length n, the expected number of spurious keys satisfies*

$$s_n \geq \frac{|\mathcal{K}|}{|\mathcal{P}|^{nR_L}} - 1.$$

Unicity Distance

The unicity distance of a cryptosystem is the the average size of cipheretext at which the expected number of spurious keys becomes zero. Using our previous theorem, we get an estimate for the unicity distance of $\frac{\log_2 |\mathcal{K}|}{R_L \log_2 |\mathcal{P}|}$.

Example. Consider the Shift Cipher . Here $|\mathcal{P}| = |\mathcal{K}| = 26$. Then the unicity distance is approximately

$$\frac{\log_2 26}{R_L \log_2 26} = \frac{1}{R_L} = 1.33.$$

Then, although the cipher is perfectly secret for one character, it is quite insecure for two.

Example. Consider the Substitution Cipher.

http://www.gutenberg.org/files/108/108−h/108−h.htm

Here $|\mathcal{P}| = 26$ and $|\mathcal{K}| = 26!$. Then we get an estimate of the unicity distance of

$$\frac{log_2 26!}{R_L log_2 26} = \frac{88.4}{0.75 \times 4.7} = 25.$$

Then, given a ciphertext of length at least 25, (usually) a unique decryption is possible.

Example. Consider the Jefferson wheel, consisting of 36 disks, each consisting of the 26 letters of the English alphabet in a randomized order.

Source: http://www.geocaching.com/geocache/GC23BTT_crypto−4

Then $|\mathcal{P}| = 26^{36}$ and $|\mathcal{K}| = 26^{36} \cdot (36)!$. Then we get an estimate of the unicity distance of

$$\frac{\log_2 26^{36} \cdot (36)!}{R_L \log_2 26^{36}} = \frac{136.54 + 169.2}{0.75 \cdot 169.2} \approx 2.4.$$

http://franceshunter.files.wordpress.com/2011/09/wheelcipher_lg.jpg

Acknowledgements

Stinson, Douglas. *Cryptography Theory and Practice*. Taylor & Francis Group, 2006.