COMP 8920: Cryptography Winter 2025

Lecture 5 — Jan 21, 2025

Prof. Curtis Bright *Scribe: Aidan Bennett*

Overview

In the last lecture we looked at cryptanalysis for several of the cryptosystems we had already discussed.

In this lecture we will finish our discussion of cryptanalysis of these cryptosystems, and introduce some probability concepts, as well as the notion of perfect security. This corresponds to sections 2.2.4 through 3.3 in the textbook.

Cryptanalysis of Hill Cipher (continued)

Example: Suppose the plaintext *friday* is encrypted using a Hill Cipher with $m = 2$, yielding *PQCFKU*. Then we have that

$$e_k(5, 17) = (15, 16)$$
$$e_k(8, 3) = (2, 5)$$
$$e_k(0, 24) = (10, 20)$$

Taking the first two pairs we get the matrix equation

$$\begin{bmatrix} 15 & 16 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 5 & 17 \\ 8 & 3 \end{bmatrix} K$$

We can solve for K by computing the inverse of the matrix on the right as follows.

$$\begin{bmatrix} 5 & 17 \\ 8 & 3 \end{bmatrix}^{-1} = 9^{-1} \begin{bmatrix} 3 & -17 \\ -8 & 5 \end{bmatrix} = 3 \begin{bmatrix} 3 & -17 \\ -8 & 5 \end{bmatrix} = \begin{bmatrix} 9 & 1 \\ 2 & 15 \end{bmatrix}$$

Multiplying by this on both sides of the equation, we find

$$K = \begin{bmatrix} 9 & 1 \\ 2 & 15 \end{bmatrix} \begin{bmatrix} 15 & 16 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 17 & 19 \\ 8 & 3 \end{bmatrix}$$

which can be easily verified using the remaining known plaintext and ciphertext.

Cryptanalysis of LFSR stream cipher

Recall that for an LFSR stream cipher, the keystream is generated from initial values $(z_1, \ldots, z_m) = (k_1, \ldots, k_m)$, together with a linear recurrence

$$z_{i+m} = \sum_{j=0}^{m-1} c_j z_{i+j} \mod 2 \tag{1}$$

for $i \geq 1$, and fixed coefficients $c_0, \ldots, c_{m-1} \in \mathbb{Z}_2$.

We will use a **known plaintext attack**, assuming m is known. Suppose (x, y) is a known plaintext/ciphertext pair. We will denote $n := \text{len}(x)$, and assume $n \geq 2m$. Recall that y is generated from x by a bitwise XOR with the keystream, so $y_i = x_i \oplus z_i$, and thus $z_i = y_i \oplus x_i$.

So XORing x and y give the initial values $(z_1, \ldots, z_m)$, but we still need to find $c_0, \ldots, c_{m-1}$. Since Equation (1) has m unknowns, as long as we have that $n \geq 2m$, we can create a system of linear equations that we can solve:

$$(z_{m+1}, \ldots, z_{2m}) = (c_0, \ldots, c_{m-1}) \underbrace{\begin{pmatrix} z_1 & z_2 & \cdots & z_m \\ z_2 & z_3 & \cdots & z_{m+1} \\ \vdots & \vdots & & \vdots \\ z_m & z_{m+1} & \cdots & z_{2m-1} \end{pmatrix}}_{\text{(denoted } Z\text{)}}$$

We can then compute $(c_0, \ldots, c_{m-1}) = (z_{m+1}, \ldots, z_{2m})Z^{-1}$. Note that Z will always be invertible if the recurrence used to generate the keystream has degree m.

Example: Suppose we have the known plaintext/ciphertext pair for an LFSR stream cipher where $x = 0011$, and $y = 1110$, and m is known to be 2. We can compute $z = x \oplus y = (1, 1, 0, 1)$. Then

$$(z_3, z_4) = (c_0, c_1) \begin{pmatrix} z_1 & z_2 \\ z_2 & z_3 \end{pmatrix}$$

$$(c_0, c_1) = (0, 1) \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{-1} = (1, 1)$$

So $z_{i+2} = z_{i+1} + z_i$, like the Fibonacci recurrence.

Notions of security

Recall the notions of security:

1. **Computational security**: Security of cryptosystem relies on being safe against the current best known attacks; safe as long as no new attack methods are discovered
2. **Provable security:** It can be proved that the problem of breaking a cryptosystem reduces to a problem that is known to be difficult, like factoring. Safe as long as related more difficult problem remains unsolved.
3. **Unconditional security**: Impossible to break, even with infinite time and resources. We will be able to use probability to show when a cryptosystem is unconditionally secure.

Probability Theory

Definition 1. *A **discrete random variable** $\mathbf{X}$ consists of a finite set X, and a **probability distribution** defined on X. The probability that a random variable $\mathbf{X}$ takes on the value x is denoted $\Pr[\mathbf{X} = x]$ (or just $\Pr[x]$ if the random variable is clear).*

Some more important information:

- The axioms of probability state that $0 \leq \Pr[x]$ for all x, and
$$\sum_{x \in X} \Pr[x] = 1$$
- The *joint probability distribution* of two random variables $\mathbf{X}, \mathbf{Y}$, denoted $\Pr[\mathbf{X} = x, \mathbf{Y} = y]$, is the probability that $\mathbf{X}$ takes the value x while also $\mathbf{Y}$ takes the value y.
- A **conditional probability**, $\Pr[x|y]$, is the probability that $\mathbf{X}$ takes on the value x *given* that $\mathbf{Y}$ takes on the value y.
- Two random variables $\mathbf{X}$ and $\mathbf{Y}$ are called **independent** if $\Pr[x, y] = \Pr[x]\Pr[y]$, for all x, y.

Some important formulas relating conditional probabilities to joint probabilities include:

- $\Pr[x, y] = \Pr[x|y]\Pr[y]$
- By symmetry, we also get $\Pr[x, y] = \Pr[y|x]\Pr[x]$
- Combining the above two, we get Bayes' Theorem:
$$\Pr[x|y] = \frac{\Pr[y|x]\Pr[x]}{\Pr[y]}$$

Perfect Security

Let $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ be a cryptosystem. We will assume the key $k \in \mathcal{K}$ is always only used once. Suppose there is a probability distribution on the plaintext space $\mathcal{P}$, such that the plaintext defines a random variable $\mathbf{X}$, and $\Pr[x]$ is the probability that message x is sent.

Similarly, $\Pr[\mathbb{K} = k]$ is the probability of k being the chosen key. We will assume that $\mathbb{K}, \mathbf{X}$ are independent. These random variables induce a probability distribution on $\mathcal{C}$, where $\Pr[y]$ is the probability of the message $y \in \mathcal{C}$ being sent.

For a key $k \in \mathcal{K}$, define $C(k) := \{e_k(x) : x \in \mathcal{P}\}$ to be the set of all possible ciphertexts y that k could encrypt to. Then
$$\Pr[y] = \sum_{\{k : y \in C(k)\}} \Pr[\mathbb{K} = k]\Pr[\mathbf{X} = d_k(y)]$$
It also follows that
$$\Pr[y|x] = \sum_{\{k : y = e_k(x)\}} \Pr[\mathbb{K} = k]$$
Combining the above two formulas with Baye's Theorem yields
$$\Pr[x|y] = \frac{\Pr[x] \sum_{\{k : y = e_k(x)\}} \Pr[\mathbb{K} = k]}{\sum_{\{k : y \in C(k)\}} \Pr[\mathbb{K} = k]\Pr[\mathbf{X} = d_k(y)]}$$

	a	c
k_1	1	2
k_2	2	3
k_3	3	4

Table 1: Encryption matrix

Example: Consider the cryptosystem where $\mathcal{P} = \{a, b\}$, $\mathcal{C} = \{1, 2, 3, 4\}$, $\mathcal{K} = \{k_1, k_2, k_3\}$, and encryption is done according to Table 1. Assume that $\Pr[a] = \frac{1}{4}$, $\Pr[b] = \frac{3}{4}$, $\Pr[k_1] = \frac{1}{2}$, $\Pr[k_2] = \frac{1}{4}$, $\Pr[k_3] = \frac{1}{4}$.

Then

$$\Pr[1] = \Pr[k_1]\Pr[\mathbf{X} = d_{k_1}(1)] = \frac{1}{2} \cdot \frac{1}{4} = \frac{7}{16}$$

$$\Pr[2] = \Pr[k_1]\Pr[\mathbf{X} = d_{k_2}(2)] + \Pr[k_2]\Pr[\mathbf{X} = d_{k_2}(2)] = \frac{1}{2} \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{1}{4} = \frac{7}{16}$$

$$\Pr[a|1] = \frac{\Pr[a] \sum_{k \in \{k_1\}} \Pr[\mathbb{K} = k]}{\Pr[1]} = \frac{\frac{1}{4} \cdot \frac{1}{2}}{\frac{1}{8}} = 1$$

A cryptosystem being **perfectly secure** means that when the ciphertext is observed, no information is revealed about the plaintext. Formally, for all $(x, y) \in \mathcal{P} \times \mathcal{C}$, $\Pr[x|y] = \Pr[x]$.