Complexity and Cryptography

Thomas Zeugmann

Hokkaido University
Laboratory for Algorithmics

https://www-alg.ist.hokudai.ac.jp/~thomas/COCR/

Lecture 12: Security and One-Time-Pads

Introduction I

We aim to study the security of cryptosystems from a higher point of view. Instead of looking at a particular cryptosystem, we are interested in general properties a cryptosystems must possess to be secure. We shall distinguish between

- *unconditionally secure*,
- *computationally secure*,
- *provably secure*, and
- *insecure*

cryptosystems.

Introduction II

Intuitively, a cryptosystem is said to be *unconditionally secure* if the probability p, $p < 1$, of breaking it, is independent of the computing resources available and of the time at an adversary's disposal.

Introduction II

Intuitively, a cryptosystem is said to be *unconditionally secure* if the probability p, $p < 1$, of breaking it, is independent of the computing resources available and of the time at an adversary's disposal.

We call a cryptosystem *computationally secure* if braking it is possible in principle but all *known* methods of executing the computation necessary require an infeasible amount of time and/or hardware.

Introduction II

Intuitively, a cryptosystem is said to be *unconditionally secure* if the probability p, $p < 1$, of breaking it, is independent of the computing resources available and of the time at an adversary's disposal.

We call a cryptosystem *computationally secure* if braking it is possible in principle but all *known* methods of executing the computation necessary require an infeasible amount of time and/or hardware.

A cryptosystem is said to be *provably secure* if it can be shown that breaking it for any *significant* number of cases implies that some other problem – such as computing the factorization of large composite integers – could be solved with comparable effort.

Introduction III

The distinction between computationally secure and provably secure is that while in either case the security of the system would be impeached if the underlying computationally difficult problem could be solved, the converse does not need to hold for computationally secure systems but *does hold* for provable secure systems. To provide these intuitive definitions was easy, but it requires extra work to make them mathematically sound.

We start with the notion of *unconditionally secure cryptosystems*.

Unconditionally Secure Cryptosystems I

We model cryptosystems as follows: By $\mathbb{Z}_m$ we denote any fixed alphabet of size m and we set $\mathbb{Z}_{m,n} = \mathbb{Z}_m^n$. The elements of $\mathbb{Z}_{m,n}$ are called n-*grams*. By Pt and Ct we denote the set of all possible plaintexts and ciphertexts, respectively, i.e., we define $\mathrm{Pt} =_{df} \mathrm{Ct} =_{df} \bigcup_{n\in\mathbb{N}} \mathbb{Z}_{m,n}$.

Unconditionally Secure Cryptosystems I

We model cryptosystems as follows: By $\mathbb{Z}_m$ we denote any fixed alphabet of size m and we set $\mathbb{Z}_{m,n} = \mathbb{Z}_m^n$. The elements of $\mathbb{Z}_{m,n}$ are called n-*grams*. By Pt and Ct we denote the set of all possible plaintexts and ciphertexts, respectively, i.e., we define $\mathrm{Pt} =_{df} \mathrm{Ct} =_{df} \bigcup_{n\in\mathbb{N}} \mathbb{Z}_{m,n}$.

Definition 1

A *cryptographic transformation* T is a sequence of bijective transformations $(T^{(n)})_{n\in\mathbb{N}}$ with $T^{(n)} : \mathbb{Z}_{m,n} \to \mathbb{Z}_{m,n}$, where the set $\mathbb{Z}_m$ is arbitrarily fixed.
A *cryptosystem* $\mathcal{T}$ is a family of cryptographic transformations, i.e., $\mathcal{T} =_{df} \{T_k \mid k \in K\}$. We refer to K as the set of all admissible *keys* (admissible means admissible for $\mathcal{T}$), and every $k \in K$ is called a *key*.

Unconditionally Secure Cryptosystems II

In accordance with Lecture 11, we generally assume that $\mathcal{T}$ is known to the cryptanalyst but she does not know which key k; i.e., which transformation $T_k^{(n)}$, has been used. Furthermore, we assume that the generation of plaintext and the choice of a key are independent probabilistic processes.

Unconditionally Secure Cryptosystems II

In accordance with Lecture 11, we generally assume that $\mathcal{T}$ is known to the cryptanalyst but she does not know which key k; i.e., which transformation $T_k^{(n)}$, has been used. Furthermore, we assume that the generation of plaintext and the choice of a key are independent probabilistic processes.

The cryptanalyst has to determine the plaintext and the key, respectively, by using the available information derivable from the following:

Unconditionally Secure Cryptosystems III

1. the ciphertext Y received,

Unconditionally Secure Cryptosystems III

1. the ciphertext Y received,
2. the cryptosystem $\mathcal{T} = \{T_k \mid k \in K\}$ used to compute Y,

Unconditionally Secure Cryptosystems III

1. the ciphertext Y received,
2. the cryptosystem $\mathcal{T} = \{T_k \mid k \in K\}$ used to compute Y,
3. *a priori* assumptions about the probability distribution Pr_{plain} over the set of all plaintexts,

Unconditionally Secure Cryptosystems III

1. the ciphertext Y received,
2. the cryptosystem $\mathcal{T} = \{T_k \mid k \in K\}$ used to compute Y,
3. *a priori* assumptions about the probability distribution Pr_{plain} over the set of all plaintexts,
4. *a priori* assumptions about the probability distribution Pr_{key} over the set K of all keys admissible for $\mathcal{T}$,

Unconditionally Secure Cryptosystems III

1. the ciphertext Y received,
2. the cryptosystem $\mathcal{T} = \{T_k \mid k \in K\}$ used to compute Y,
3. *a priori* assumptions about the probability distribution Pr_{plain} over the set of all plaintexts,
4. *a priori* assumptions about the probability distribution Pr_{key} over the set K of all keys admissible for $\mathcal{T}$,
5. some possible ciphers $y = T_k(x)$, where $k \in K$, $x \in Pt$.

Unconditionally Secure Cryptosystems III

1. the ciphertext Y received,
2. the cryptosystem $\mathcal{T} = \{T_k \mid k \in K\}$ used to compute Y,
3. *a priori* assumptions about the probability distribution Pr_{plain} over the set of all plaintexts,
4. *a priori* assumptions about the probability distribution Pr_{key} over the set K of all keys admissible for $\mathcal{T}$,
5. some possible ciphers $y = T_k(x)$, where $k \in K$, $x \in Pt$.

The *a priori* assumptions about Pr_{plain} and Pr_{key} induce a probability distribution Pr_{cipher} over the set of all ciphertexts Ct. Additionally, the following probabilities can be determined:

Unconditionally Secure Cryptosystems IV

(1) $Pr_{plain,key}(x,k) =_{df} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability that for deciphering x the key k has been used.

Unconditionally Secure Cryptosystems IV

(1) $Pr_{plain,key}(x,k) =_{df} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability that for deciphering x the key k has been used.

(2) $Pr_{cipher}(y) =_{df} \sum_{\{(x,k)|T_k(x)=y\}} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability of the ciphertext y to occur.

Unconditionally Secure Cryptosystems IV

(1) $Pr_{plain,key}(x,k) =_{df} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability that for deciphering x the key k has been used.

(2) $Pr_{cipher}(y) =_{df} \sum_{\{(x,k)|T_k(x)=y\}} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability of the ciphertext y to occur.

(3) $Pr_{plain,cipher}(x,y) =_{df} \sum_{\{k\in K|T_k(x)=y\}} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability that y is the enciphering of the plaintext x.

Unconditionally Secure Cryptosystems IV

(1) $Pr_{plain,key}(x,k) =_{df} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability that for deciphering x the key k has been used.

(2) $Pr_{cipher}(y) =_{df} \sum_{\{(x,k)|T_k(x)=y\}} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability of the ciphertext y to occur.

(3) $Pr_{plain,cipher}(x,y) =_{df} \sum_{\{k\in K|T_k(x)=y\}} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability that y is the enciphering of the plaintext x.

(4) $Pr_{cipher,\,plain}(y,x) =_{df} \sum_{\{k\in K|T_k^{-1}(y)=x\}} Pr_{cipher}(y) \cdot Pr_{key}(k)$;

i.e., the probability that the cipher y originates from the plaintext x.

Unconditionally Secure Cryptosystems IV

(1) $Pr_{plain,key}(x,k) =_{df} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability that for deciphering x the key k has been used.

(2) $Pr_{cipher}(y) =_{df} \sum_{\{(x,k)|T_k(x)=y\}} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability of the ciphertext y to occur.

(3) $Pr_{plain,cipher}(x,y) =_{df} \sum_{\{k\in K|T_k(x)=y\}} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability that y is the enciphering of the plaintext x.

(4) $Pr_{cipher,plain}(y,x) =_{df} \sum_{\{k\in K|T_k^{-1}(y)=x\}} Pr_{cipher}(y) \cdot Pr_{key}(k)$; i.e., the probability that the cipher y originates from the plaintext x.

(5) $Pr_{cipher,key}(y,k) =_{df} \sum_{\{x\in Pt|T_k(x)=y\}} Pr_{plain}(x) \cdot Pr_{key}(k)$; i.e., the probability that that for cipher y the key k was used.

Unconditionally Secure Cryptosystems V

(6) $\Pr_{\text{plain}|\text{cipher}}(x|y) =_{df} \frac{\Pr_{\text{plain, cipher}}(x,y)}{\Pr_{\text{cipher}}(y)}$; i.e., the conditional probability of the plaintext x under the observation of the cipher y.

Unconditionally Secure Cryptosystems V

(6) $\mathrm{Pr}_{\mathrm{plain}|\mathrm{cipher}}(x|y) =_{\mathrm{df}} \frac{\mathrm{Pr}_{\mathrm{plain},\mathrm{cipher}}(x,y)}{\mathrm{Pr}_{\mathrm{cipher}}(y)}$; i.e., the conditional probability of the plaintext x under the observation of the cipher y.

(7) $\mathrm{Pr}_{\mathrm{key}|\mathrm{cipher}}(k|y) =_{\mathrm{df}} \frac{\mathrm{Pr}_{\mathrm{cipher},\mathrm{key}}(y,k)}{\mathrm{Pr}_{\mathrm{cipher}}(y)}$; i.e., the conditional probability of the key k under the observation of the cipher y.

Unconditionally Secure Cryptosystems V

(6) $\Pr_{\text{plain}|\text{cipher}}(x|y) =_{df} \frac{\Pr_{\text{plain, cipher}}(x,y)}{\Pr_{\text{cipher}}(y)}$; i.e., the conditional probability of the plaintext x under the observation of the cipher y.

(7) $\Pr_{\text{key}|\text{cipher}}(k|y) =_{df} \frac{\Pr_{\text{cipher, key}}(y,k)}{\Pr_{\text{cipher}}(y)}$; i.e., the conditional probability of the key k under the observation of the cipher y.

(8) $\Pr_{\text{cipher}|\text{plain}}(y|x) = \frac{\Pr_{\text{cipher, plain}}(y,x)}{\Pr_{\text{plain}}(x)}$; i.e., the conditional probability of the cipher y under the observation of the plaintext x.

Unconditionally Secure Cryptosystems V

(6) $\Pr_{\text{plain}|\text{cipher}}(x|y) =_{df} \frac{\Pr_{\text{plain, cipher}}(x,y)}{\Pr_{\text{cipher}}(y)}$; i.e., the conditional probability of the plaintext x under the observation of the cipher y.

(7) $\Pr_{\text{key}|\text{cipher}}(k|y) =_{df} \frac{\Pr_{\text{cipher, key}}(y,k)}{\Pr_{\text{cipher}}(y)}$; i.e., the conditional probability of the key k under the observation of the cipher y.

(8) $\Pr_{\text{cipher}|\text{plain}}(y|x) = \frac{\Pr_{\text{cipher, plain}}(y,x)}{\Pr_{\text{plain}}(x)}$; i.e., the conditional probability of the cipher y under the observation of the plaintext x.

For the definition of the conditional probabilities we must assume that $\Pr_{\text{cipher}}(y) > 0$ and $\Pr_{\text{plain}}(x) > 0$, respectively. If these probabilities are zero, we define the respective conditional probabilities to be zero, too.

Unconditionally Secure Cryptosystems VI

Now, we are in a position to define the notion of unconditional security.

Definition 2

A cryptosystem $\mathcal{T}$ is *unconditionally secure* if

$$Pr_{plain|cipher}(x|y) = Pr_{plain}(x)$$

for all plaintexts x and ciphertexts y with $Pr_{cipher}(y) > 0$.

Unconditionally Secure Cryptosystems VI

Now, we are in a position to define the notion of unconditional security.

Definition 2

A cryptosystem $\mathcal{T}$ is *unconditionally secure* if

$$\Pr_{\text{plain}|\text{cipher}}(x|y) = \Pr_{\text{plain}}(x)$$

for all plaintexts x and ciphertexts y with $\Pr_{\text{cipher}}(y) > 0$.

Thus, for an unconditionally secure cryptosystem the probability distributions $\Pr_{\text{plain}}$ and $\Pr_{\text{plain}|\text{cipher}}(\,\cdot\,|y)$ are identical for all ciphers y provided the cipher can be generated at all by the cryptosystem $\mathcal{T}$. In other words, whatever the eavesdropped cipher y is, for the cryptanalyst the probability to break it is the same as having not seen it.

Unconditionally Secure Cryptosystems VII

Note that Definition 2 does not make any assumptions concerning the computing power available to the cryptanalyst nor does it limit the time she may spend.

Unconditionally Secure Cryptosystems VII

Note that Definition 2 does not make any assumptions concerning the computing power available to the cryptanalyst nor does it limit the time she may spend.

The following exercise provides a first characterization of unconditionally secure cryptosystems:

Exercise 1. *A cryptosystem $\mathcal{T}$ is unconditionally secure if and only if $Pr_{cipher|plain}(y|x) = Pr_{cipher}(y)$ for all plaintexts x with $Pr_{plain}(x) > 0$.*

Unconditionally Secure Cryptosystems VII

Note that Definition 2 does not make any assumptions concerning the computing power available to the cryptanalyst nor does it limit the time she may spend.

The following exercise provides a first characterization of unconditionally secure cryptosystems:

Exercise 1. *A cryptosystem $\mathcal{T}$ is unconditionally secure if and only if $\Pr_{\text{cipher}|\text{plain}}(y|x) = \Pr_{\text{cipher}}(y)$ for all plaintexts x with $\Pr_{\text{plain}}(x) > 0$.*

The following theorem provides a necessary condition for unconditionally secure cryptosystems by relating the number of keys necessary to the number of plaintexts having non-zero probability:

Unconditionally Secure Cryptosystems VIII

Theorem 1

Let $\mathcal{T}$ be any cryptosystem that is unconditionally secure for all n-grams x, y with $\Pr_{\text{plain}}(x) > 0$ and $\Pr_{\text{cipher}}(y) > 0$. Then

$$|K| \geqslant \left|\left\{x \in \mathbb{Z}_{m,n} \mid \Pr_{\text{plain}}(x) > 0\right\}\right| .$$

Unconditionally Secure Cryptosystems VIII

Theorem 1

Let $\mathcal{T}$ be any cryptosystem that is unconditionally secure for all n-grams x, y with $\Pr_{\text{plain}}(x) > 0$ and $\Pr_{\text{cipher}}(y) > 0$. Then

$$|K| \geqslant \left|\left\{x \in \mathbb{Z}_{m,n} \mid \Pr_{\text{plain}}(x) > 0\right\}\right| .$$

Proof. We introduce the following notations:

$$\begin{aligned} \mathbb{Z}_{m,n,+,\mathit{plain}} &=_{df} \left\{x \in \mathbb{Z}_{m,n} \mid \Pr_{\text{plain}}(x) > 0\right\} , \\ \mathbb{Z}_{m,n,+,\mathit{cipher}} &=_{df} \left\{y \in \mathbb{Z}_{m,n} \mid \Pr_{\text{cipher}}(y) > 0\right\} . \end{aligned}$$

Furthermore, without loss of generality we may assume $\Pr_{\text{key}}(k) > 0$ for all $k \in K$. Otherwise, we replace K by $\widehat{K}$, where $\widehat{K}$ is the set of all keys having non-zero probability.
We need the following observations:

Unconditionally Secure Cryptosystems IX

Observation 1

For all keys $k \in K$ *the (restricted) transformation* $T_k : \mathbb{Z}_{m,n,+,plain} \to \mathbb{Z}_{m,n,+,cipher}$ *is injective.*

Unconditionally Secure Cryptosystems IX

Observation 1

For all keys $k \in K$ the (restricted) transformation $T_k : \mathbb{Z}_{m,n,+,plain} \to \mathbb{Z}_{m,n,+,cipher}$ is injective.

By definition $T_k : \mathbb{Z}_{m,n} \to \mathbb{Z}_{m,n}$ is bijective. We know that

$$Pr_{cipher}(y) = \sum_{\{(x,k)|T_k(x)=y\}} Pr_{plain}(x) \cdot Pr_{key}(k).$$

Thus, $Pr_{cipher}(y) > 0$ iff there is at least one pair (x, k) with $T_k(x) = y$ satisfying $Pr_{plain}(x) > 0$ and $Pr_{key}(k) > 0$.
Since $Pr_{key}(k) > 0$ for all $k \in K$, we directly obtain that $T_k(\mathbb{Z}_{m,n,+,plain}) \subseteq \mathbb{Z}_{m,n,+,cipher}$ for all $k \in K$. So the restriction of T_k to $\mathbb{Z}_{m,n,+,plain}$ is an injective mapping into $\mathbb{Z}_{m,n,+,cipher}$.

Unconditionally Secure Cryptosystems X

Observation 2

Let $y \in \mathbb{Z}_{m,n,+,cipher}$ *be arbitrarily fixed. Then, for every* $x \in \mathbb{Z}_{m,n,+,plain}$ *there must be a key* $k \in K$ *such that* $T_k(x) = y$.

Unconditionally Secure Cryptosystems X

Observation 2

Let $y \in \mathbb{Z}_{m,n,+,cipher}$ be arbitrarily fixed. Then, for every $x \in \mathbb{Z}_{m,n,+,plain}$ there must be a key $k \in K$ such that $T_k(x) = y$.

Suppose the converse; i.e., there exists an $\hat{x} \in \mathbb{Z}_{m,n,+,plain}$ such that $T_k(\hat{x}) \neq y$ for all $k \in K$. Hence $\{k \in K \mid T_k(\hat{x}) = y\} = \emptyset$. Since the cryptosystem is unconditionally secure we have

$$\Pr_{plain}(\hat{x}) = \Pr_{plain|cipher}(\hat{x}|y) = \frac{\Pr_{plain,\,cipher}(\hat{x}, y)}{\Pr_{cipher}(y)}$$

$$= \frac{\sum_{\{k \in K \mid T_k(\hat{x}) = y\}} \Pr_{plain}(\hat{x}) \cdot \Pr_{key}(k)}{\Pr_{cipher}(y)} = 0\,,$$

a contradiction to $\hat{x} \in \mathbb{Z}_{m,n,+,plain}$.

Unconditionally Secure Cryptosystems XI

Finally, for all $x_1, x_2 \in \mathbb{Z}_{m,n,+,plain}$ with $x_1 \neq x_2$ and all $k_1, k_2 \in K$ satisfying $T_{k_1}(x_1) = y = T_{k_2}(x_2)$ we can conclude $k_1 \neq k_2$.

In order to see this, suppose the converse; i.e., there are $x_1, x_2 \in \mathbb{Z}_{m,n,+,plain}$ with $x_1 \neq x_2$ and a key k such that $T_k(x_1) = y = T_k(x_2)$. However, T_k is injective, and therefore this would imply $x_1 = x_2$, a contradiction.

Unconditionally Secure Cryptosystems XI

Finally, for all $x_1, x_2 \in \mathbb{Z}_{m,n,+,plain}$ with $x_1 \neq x_2$ and all $k_1, k_2 \in K$ satisfying $T_{k_1}(x_1) = y = T_{k_2}(x_2)$ we can conclude $k_1 \neq k_2$.

In order to see this, suppose the converse; i.e., there are $x_1, x_2 \in \mathbb{Z}_{m,n,+,plain}$ with $x_1 \neq x_2$ and a key k such that $T_k(x_1) = y = T_k(x_2)$. However, T_k is injective, and therefore this would imply $x_1 = x_2$, a contradiction.

Consequently, the set K must contain at least as many keys as there are elements in $\mathbb{Z}_{m,n,+,plain}$. ∎

One-Time-Pads I

Next, we show the existence of unconditionally secure cryptosystems. We define a cryptosystem, the so-called *one-time-pads*, and prove it to be unconditionally secure.

One-Time-Pads I

Next, we show the existence of unconditionally secure cryptosystems. We define a cryptosystem, the so-called *one-time-pads*, and prove it to be unconditionally secure.

One-time-pads were introduced by Gilbert S. Vernam in 1918. His idea was to introduce uncertainty at the same rate at which is was removed by redundancy among symbols of the message. His intuition was right as proved more than two decades later by Claude E. Shannon (1949).

One-Time-Pads I

Next, we show the existence of unconditionally secure cryptosystems. We define a cryptosystem, the so-called *one-time-pads*, and prove it to be unconditionally secure.

One-time-pads were introduced by Gilbert S. Vernam in 1918. His idea was to introduce uncertainty at the same rate at which is was removed by redundancy among symbols of the message. His intuition was right as proved more than two decades later by Claude E. Shannon (1949).

But as Theorem 1 shows, this ideal requires exchanging an amount of keys in advance of communication that is in most cases impractical if not infeasible.

One-Time-Pads II

For the sake of convenience, we identify the possible keys of the cryptosystem to be defined by random variables. Let $\{k_i \mid 0 \leqslant i < n\}$ be independently and identically distributed random variables taking the values from $\mathbb{Z}_m$ equally likely, i.e., $\Pr(k_i = x) = 1/m$ for all $x \in \mathbb{Z}_m$, and all $i = 0, \ldots, n-1$.

One-Time-Pads II

For the sake of convenience, we identify the possible keys of the cryptosystem to be defined by random variables. Let $\{k_i \mid 0 \leqslant i < n\}$ be independently and identically distributed random variables taking the values from $\mathbb{Z}_m$ equally likely, i.e., $\Pr(k_i = x) = 1/m$ for all $x \in \mathbb{Z}_m$, and all $i = 0, \ldots, n-1$. One-time-pads are defined as bijections as follows:

$$T^{(n)} : X = (x_0, \ldots, x_{n-1}) \to Y = (y_0, \ldots, y_{n-1}) \tag{1}$$

with a randomly generated key $(k_0, \ldots, k_{n-1})$, where for all $i = 0, \ldots, n-1$

$$y_i = T^{(n)}_{k_i}(x_i) =_{df} (k_i + x_i) \mod m \tag{2}$$

must be satisfied.

One-Time-Pads III

Thus, we have $\Pr_{\text{key}}((k_0, \ldots, k_{n-1})) = 1/m^n$. Moreover, the set of all keys is $K = \mathbb{Z}_{m,n}$, and thus also $|K| = m^n$. So there are at least as many keys for enciphering the n-grams from $\mathbb{Z}_{m,n}$ as there are elements $X \in \mathbb{Z}_{m,n}$ with $\Pr_{\text{plain}}(x) > 0$. Therefore, the condition of Theorem 1 is fulfilled. Thus, for the one-time-pad we obtain the following theorem:

Theorem 2

For every plaintext source the random variables $Y_0, \ldots, Y_{n-1}$ defined by (2) are independent and identically distributed, and every Y_i, $i = 0, \ldots, n-1$, is equally distributed over $\mathbb{Z}_m$, i.e.,

$$\Pr(Y_0, \ldots, Y_{n-1}) = \frac{1}{m^n} \quad \text{for all } y = (y_0, \ldots, y_{n-1}) \in \mathbb{Z}_{m,n} .$$

One-Time-Pads IV

Proof. The values x_i and y_i uniquely determine k_i by $y_i - x_i \equiv k_i \mod m$. All k_i are independent and identically distributed. Each k_i is equally likely chosen from $\mathbb{Z}_m$. Hence,

$$\begin{aligned} \Pr_{\text{plain, cipher}}\{X = x, Y = y\} &= \sum_{\{k \in K \mid T_k^{(n)}(x) = y\}} \Pr_{\text{plain}}\{X = x\} \Pr_{\text{key}}(k) \\ &= \frac{1}{m^n} \Pr_{\text{plain}}\{X = x\} . \end{aligned}$$

One-Time-Pads V

Finally, in accordance with the definition of $\Pr_{\text{cipher}}$ we have

$$\begin{aligned}
\Pr_{\text{cipher}}\{Y = y\} &= \sum_{(x_0,\ldots,x_{n-1})\in\mathbb{Z}_{m,n}} \Pr_{\text{plain, cipher}}\{X = x, Y = y\} \\
&= \frac{1}{m^n} \sum_{(x_0,\ldots,x_{n-1})\in\mathbb{Z}_{m,n}} \Pr_{\text{plain}}\{X = x\} \\
&= \frac{1}{m^n} .
\end{aligned}$$

One-Time-Pads V

Finally, in accordance with the definition of $\Pr_{\text{cipher}}$ we have

$$\begin{aligned}
\Pr_{\text{cipher}}\{Y = y\} &= \sum_{(x_0,\ldots,x_{n-1})\in\mathbb{Z}_{m,n}} \Pr_{\text{plain, cipher}}\{X = x, Y = y\} \\
&= \frac{1}{m^n} \sum_{(x_0,\ldots,x_{n-1})\in\mathbb{Z}_{m,n}} \Pr_{\text{plain}}\{X = x\} \\
&= \frac{1}{m^n} .
\end{aligned}$$

Analogously, we can show for every position y_i in the ciphertext that $\Pr_{\text{cipher}}(y_i) = 1/m$. Therefore, the y_i are independent and identically distributed, and

$$\Pr(y_0,\ldots,y_{n-1}) = \frac{1}{m^n} \quad \text{for all } y = (y_0,\ldots,y_{n-1}) \in \mathbb{Z}_{m,n},$$

and the Theorem is shown. ∎

One-Time-Pads VI

Corollary 1

One-time pads defined by (2) *are unconditionally secure for all plaintexts of length* n.

One-Time-Pads VI

Corollary 1

One-time pads defined by (2) are unconditionally secure for all plaintexts of length n.

Proof. We compute

$$\begin{aligned} \Pr_{\text{plain}|\text{cipher}}\{X = x | Y = y\} &= \frac{\Pr_{\text{plain, cipher}}\{X = x, Y = y\}}{\Pr_{\text{cipher}}\{Y = y\}} \\ &= \frac{\frac{1}{m^n} \cdot \Pr_{\text{plain}}\{X = x\}}{\frac{1}{m^n}} \qquad (3) \\ &= \Pr_{\text{plain}}\{X = x\}, \end{aligned}$$

where Equality (3) is obtained by Theorem 2. ∎

Making A Priori Assumptions I

Until now, we left it open in which way *a priori* assumptions concerning $\Pr_{\text{plain}}$ and $\Pr_{\text{key}}$ are made. Usually, the keys are uniformly distributed, i.e., each key is equally likely. The harder part is to find appropriate models for natural languages allowing reasonable assumptions over $\Pr_{\text{plain}}$. We describe some of the possible models, and outline generalizations.

Making A Priori Assumptions I

Until now, we left it open in which way *a priori* assumptions concerning $\Pr_{\text{plain}}$ and $\Pr_{\text{key}}$ are made. Usually, the keys are uniformly distributed, i.e., each key is equally likely. The harder part is to find appropriate models for natural languages allowing reasonable assumptions over $\Pr_{\text{plain}}$. We describe some of the possible models, and outline generalizations.

First of all, we generally assume the cryptanalyst to know in which language the plaintext is written. One could try to list all possible n-grams in the relevant language, where n corresponds to the length of the ciphertext eavesdropped. However, this approach would require a too huge amount of data to be processed, and the resulting probabilities are hard to handle numerically.

Making A Priori Assumptions II

So it is common to model languages as probabilistic processes. The resulting model should fulfill the following properties:

1. The model should reflect characteristic properties of the language modeled with "sufficient" precision. For example, in German, English and French, the letter Q is always followed by a U (e.g., Quark, question, informatique), while any other combination like QE, QP, QR, never appears.

Making A Priori Assumptions II

So it is common to model languages as probabilistic processes. The resulting model should fulfill the following properties:

1. The model should reflect characteristic properties of the language modeled with "sufficient" precision. For example, in German, English and French, the letter Q is always followed by a U (e.g., Quark, question, informatique), while any other combination like QE, QP, QR, never appears.
2. It must be possible to perform a great amount of computations within the model using a reasonable amount of time and hardware.

Making A Priori Assumptions III

In principle, each language can be modeled with any desired precision. However, the complexity of the resulting models rapidly increases with the degree of precision obtained. Therefore, for ensuring the applicability of the models, one has to compromise.

Making A Priori Assumptions III

In principle, each language can be modeled with any desired precision. However, the complexity of the resulting models rapidly increases with the degree of precision obtained. Therefore, for ensuring the applicability of the models, one has to compromise.

The *basic* idea of modeling languages is as follows: A plaintext source for texts over $\mathbb{Z}_m$ is formalized as probabilistic process; i.e., as a finite or infinite sequence of random variables $X_0, X_1, \ldots$. That is, the source models the generation of a plaintext by a random experiment resulting in a sequence of letters $x_0, x_1, \ldots$.

Making A Priori Assumptions IV

A source is defined by determining the probabilities

$$\Pr_{\text{plain}}\{X_j = x_0, X_{j+1} = x_1, \ldots, X_{j+n-1} = x_{n-1}\} \tag{4}$$

for every n-gram $(x_0, \ldots, x_{n-1}) \in \mathbb{Z}_{m,n}$ and all $j, n \in \mathbb{N}$.
To obtain mathematically sound models, the entity of all defined n-gram probabilities $\Pr_{\text{plain}}(x_0, \ldots, x_{n-1})$ must fulfill the following conditions:

Making A Priori Assumptions V

(1) $\Pr_{\text{plain}}(x_0, \ldots, x_{n-1}) \geqslant 0$ for all $n \in \mathbb{N}$ and all $(x_0, \ldots, x_{n-1}) \in \mathbb{Z}_{m,n}$,

(2) $$\sum_{(x_0,\ldots,x_{n-1}) \in \mathbb{Z}_{m,n}} \Pr_{\text{plain}}(x_0, \ldots, x_{n-1}) = 1,$$

(3) $$\Pr_{\text{plain}}(x_0, \ldots, x_{n-1}) = \sum_{(x_n,\ldots,x_{s-1}) \in \mathbb{Z}_{m,s-n}} \Pr_{\text{plain}}(x_0, \ldots, x_{s-1}) \quad \text{for all } s > n.$$

Making A Priori Assumptions V

(1) $\Pr_{\text{plain}}(x_0, \ldots, x_{n-1}) \geqslant 0$ for all $n \in \mathbb{N}$ and all $(x_0, \ldots, x_{n-1}) \in \mathbb{Z}_{m,n}$,

(2) $$\sum_{(x_0,\ldots,x_{n-1}) \in \mathbb{Z}_{m,n}} \Pr_{\text{plain}}(x_0, \ldots, x_{n-1}) = 1,$$

(3) $$\Pr_{\text{plain}}(x_0, \ldots, x_{n-1}) = \sum_{(x_n,\ldots,x_{s-1}) \in \mathbb{Z}_{m,s-n}} \Pr_{\text{plain}}(x_0, \ldots, x_{s-1}) \quad \text{for all } s > n.$$

Condition (1) and (2) are the classical axioms of non-negativity and normalization, respectively. Property (3) is a special case of Kolmogorov's consistency requirement. It establishes the connection between the probability of a prefix $(x_0, \ldots, x_{n-1})$ and the set of all s-grams, $s > n$, extending it. To understand its importance, it is helpful to reflect about composed words in English, e.g. furthermore, and moreover, or of composed words in German, e.g., Bügeleisen, Eisenbügel.

Variant 1: 1-gram Source

We look at different possibilities for modeling which differ from one another with respect to the degree of accuracy achieved.

Definition 3

A plaintext source generates *1-grams* over $\mathbb{Z}_m$ by identical, independent random experiments if

$Pr_{plain}(x_0, \ldots, x_{n-1}) = \prod_{i=0}^{n-1} p(x_i)$, where $p(x_i)$ denotes the probability to obtain x_i.

Variant 1: 1-gram Source

We look at different possibilities for modeling which differ from one another with respect to the degree of accuracy achieved.

Definition 3

A plaintext source generates *1-grams* over $\mathbb{Z}_m$ by identical, independent random experiments if

$\Pr_{\text{plain}}(x_0, \ldots, x_{n-1}) = \prod_{i=0}^{n-1} p(x_i)$, where $p(x_i)$ denotes the probability to obtain x_i.

Hence, we have to define p over $\mathbb{Z}_m$ such that $p(t) \geqslant 0$ for all $t \in \mathbb{Z}_m$, and $\sum_{t \in \mathbb{Z}_m} p(t) = 1$. The desired probabilities $p(t)$, $t \in \mathbb{Z}_m$ are empirically obtained by frequency analysis. But this is something we have already done for English when attacking the Vigenère cryptosystem.

Variant 1: 1-gram Source

Now, it is easy to verify that Conditions (1) and (2) above are fulfilled for the distribution $\Pr_{\text{plain}}$ given in Definition 3.

Exercise 2. *Prove that the distribution* $\Pr_{\text{plain}}$ *given in Definition 3 does satisfy Condition (3) above.*

Variant 1: 1-gram Source

Now, it is easy to verify that Conditions (1) and (2) above are fulfilled for the distribution $\Pr_{\text{plain}}$ given in Definition 3.

Exercise 2. *Prove that the distribution $\Pr_{\text{plain}}$ given in Definition 3 does satisfy Condition (3) above.*

However, modeling languages by 1-gram plaintext source is still rough. For example, $\Pr_{\text{plain}}(\text{HELP}) = \Pr_{\text{plain}}(\text{LHEP})$. Also, the characteristic mentioned above that Q must be followed by U is not reflected, since $\Pr_{\text{plain}}(\text{QE}) = 0.000156 > 0$. It may be, nevertheless, successfully applied when trying to break messages enciphered by simple cryptosystems.

Variant 2: 2-gram Source

Some of the weaknesses mentioned above are overcome by the following generalization of Definition 3:

Definition 4

A plaintext source generates *2-grams* over $\mathbb{Z}_m$ by identical and independent random experiments if

$$\Pr_{\text{plain}}(x_0, \ldots, x_{2n-1}) = \prod_{i=0}^{n-1} p(x_{2i}, x_{2i+1}) \,,$$

where $p(x_i, x_j)$ denotes the probability to obtain $x_i x_j$.

Variant 2: 2-gram Source

Hence, we have to define $p\colon \mathbb{Z}_m \times \mathbb{Z}_m \to [0,1]$ such that $p(t,s) \geqslant 0$ for all $(t,s) \in \mathbb{Z}_m \times \mathbb{Z}_m$, and $\sum_{(t,s)\in\mathbb{Z}_m\times\mathbb{Z}_m} p(t,s) = 1$.

Variant 2: 2-gram Source

Hence, we have to define $p\colon \mathbb{Z}_m \times \mathbb{Z}_m \to [0,1]$ such that $p(t,s) \geqslant 0$ for all $(t,s) \in \mathbb{Z}_m \times \mathbb{Z}_m$, and $\sum_{(t,s)\in\mathbb{Z}_m\times\mathbb{Z}_m} p(t,s) = 1$.

Again, the probability measure p is obtained by performing a frequency analysis with respect to the language in which the expected plaintexts are written. For arriving at reasonably precise estimates for the desired probabilities one has, however, to analyze much larger samples. For example, below we show the relative frequencies of all 676 possible 2-grams over $\mathbb{Z}_{26} \times \mathbb{Z}_{26}$ obtained by analyzing a sample of size 67320 of English 2-grams. The entry in the row i and column j stands for the number $N(i,j)$ of occurrences of the 2-gram (i,j) in the sample.

Variant 2: 2-gram Source

The desired probabilities are then obtained by computing

$$p(t,s) = N(t,s)/67320\,.$$

We leave it to the reader to perform this calculation. Nevertheless, a closer look to these Figures impressively shows that the characteristics of English are much better reflected by 2-gram plaintext sources than by 1-gram ones. For example, all entries that are 0 result in zero probability, too. Thus, $\Pr_{\text{plain}}(\text{QA}) = \Pr_{\text{plain}}(\text{QB}) = \ldots = \Pr_{\text{plain}}(\text{QT}) = 0$. Moreover, $\Pr_{\text{plain}}(\text{HELP}) = 0.0000061 > 0 = \Pr_{\text{plain}}(\text{LHEP})$. On the other hand, $\Pr_{\text{plain}}(\text{HELP}) = 0.0000061 < 0.0000064 = \Pr_{\text{plain}}(\text{HEPL})$, despite the fact that HEPL is not an English word.

Frequency of 2-grams, Part 1

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	7	125	251	304	13	65	151	13	311	13	67	681	182
B	114	7	2	1	394	0	0	0	74	7	0	152	6
C	319	0	52	1	453	0	0	339	202	0	86	98	4
D	158	3	4	33	572	1	20	1	273	5	0	19	27
E	492	27	323	890	326	106	93	16	118	4	27	340	253
F	98	0	0	0	150	108	0	0	188	0	0	35	1
G	122	0	0	2	271	0	20	145	95	0	0	23	3
H	646	2	5	3	2053	0	0	2	426	0	0	6	6
I	236	51	476	285	271	80	174	1	10	0	31	352	184
J	18	0	0	0	26	0	0	0	5	0	0	0	0
K	14	1	0	1	187	1	0	7	56	0	4	7	1
L	359	5	6	197	513	28	29	0	407	0	21	378	22
M	351	65	5	0	573	2	0	0	259	0	0	2	126
N	249	2	281	761	549	46	630	6	301	4	30	33	47
O	48	57	91	130	21	731	46	14	52	8	44	234	397
P	241	0	1	0	310	0	0	42	75	0	0	144	13
Q	0	0	0	0	0	0	0	0	0	0	0	0	0
R	470	15	79	129	1280	14	80	8	541	0	94	75	139
S	200	4	94	9	595	8	0	186	390	0	30	48	37
T	381	2	22	1	872	4	1	2161	865	0	0	62	27
U	72	87	103	51	91	11	80	2	54	0	3	230	69
V	65	0	0	2	522	0	0	0	223	0	0	0	1
W	282	1	0	4	239	0	0	175	259	0	0	5	0
X	9	0	15	0	17	0	0	1	15	0	0	0	1
Y	17	1	3	2	84	0	0	0	20	0	1	5	11
Z	18	0	0	0	36	0	0	0	17	0	0	1	0

Frequency of 2-grams, Part 2

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	1216	5	144	0	764	648	1019	89	137	37	17	202	15
B	0	118	0	0	81	28	6	89	2	0	0	143	0
C	3	606	0	1	113	23	237	92	0	0	0	25	0
D	8	111	0	1	49	75	2	91	15	6	0	40	0
E	1029	30	143	25	1436	917	301	36	160	153	113	90	3
F	1	326	0	0	142	3	58	54	0	0	0	5	0
G	51	129	0	0	150	29	28	58	0	0	0	6	0
H	14	287	0	0	56	10	85	31	0	4	0	15	0
I	1550	554	62	5	212	741	704	7	155	0	14	1	49
J	0	45	0	0	1	0	0	48	0	0	0	0	0
K	20	7	0	0	3	39	1	1	0	0	0	4	0
L	1	208	11	0	9	104	68	72	15	3	0	219	0
M	8	240	139	0	5	47	1	65	1	0	0	37	0
N	88	239	2	3	5	340	743	56	31	8	1	71	2
O	1232	125	164	0	861	201	223	533	188	194	7	23	2
P	1	268	103	0	409	32	51	81	0	0	0	3	0
Q	0	0	0	0	0	0	0	73	0	0	0	0	0
R	149	510	25	0	97	300	273	88	65	8	1	140	0
S	7	234	128	3	9	277	823	192	0	13	0	27	0
T	9	756	2	0	295	257	131	120	3	54	0	125	3
U	318	4	81	0	306	256	263	6	3	0	2	3	1
V	0	46	0	0	0	2	0	1	1	0	0	5	0
W	44	159	0	0	13	45	2	0	0	0	0	3	0
X	0	1	47	0	0	0	23	0	0	0	5	0	0
Y	5	64	9	0	9	44	5	4	0	3	0	2	1
Z	0	4	0	0	0	0	0	1	0	0	0	0	2

Variant 3: Markov Chains

Definition 5

A plaintext source generates *1-grams* over $\mathbb{Z}_m$ by a Markov chain with transition probabilities $P = (p(s|t))_{s,t\in\mathbb{Z}_m}$ and initial probability distribution $\pi = (\pi_0, \ldots, \pi_{m-1})$ if

$$Pr_{plain}(x_0, \ldots, x_{n-1}) = \pi(x_0)p(x_1|x_0)p(x_2|x_1)\ldots p(x_{n-1}|x_{n-2})$$

for all $n \in \mathbb{N}$ and every n-gram $(x_0, \ldots, x_{n-1}) \in \mathbb{Z}_{m,n}$.

Thereby, the following properties must be fulfilled:

Variant 3: Markov Chains

Properties

(α) $p(s|t) \geqslant 0$ for all $0 \leqslant s, t < m$,

(β) $\sum_{0 \leqslant s < m} p(s|t) = 1$ for all $0 \leqslant t < m$,

(γ) $\pi(t) \geqslant 0$ for all $t = 0, \ldots, m-1$, and $\sum_{0 \leqslant s < m} \pi(s) = 1$,

(δ) $\pi(s) = \sum_{0 \leqslant t < m} \pi(t) p(s|t)$ for all $s = 0, \ldots, m-1$.

Variant 3: Markov Chains

Properties

(α) $p(s|t) \geqslant 0$ for all $0 \leqslant s, t < m$,

(β) $\sum_{0 \leqslant s < m} p(s|t) = 1$ for all $0 \leqslant t < m$,

(γ) $\pi(t) \geqslant 0$ for all $t = 0, \ldots, m-1$, and $\sum_{0 \leqslant s < m} \pi(s) = 1$,

(δ) $\pi(s) = \sum_{0 \leqslant t < m} \pi(t) p(s|t)$ for all $s = 0, \ldots, m-1$.

This looks much more complicated than our previous definitions. Hence, some additional remarks are in order. The general idea behind the Markov chain model is to describe a language by a *probabilistic* automaton having $|\mathbb{Z}_m|$ states. Thus, each state stands for a letter.

Variant 3: Markov Chains

Intuitively, the initial probability distribution π describes the probability of a plaintext to start with a particular letter. It could be obtained by counting the number of sentences (paragraphs) in a sufficiently long text starting with the letter A, B, . . . , and Z and dividing these numbers by the number of all sentences (paragraphs) in this text.

There is, however, a better method for computing it which we describe below. Figure 1 displays these probabilities.

Variant 3: Markov Chains

letter	π	letter	π	letter	π	letter	π
A	0.0723	H	0.0402	O	0.0716	V	0.0117
B	0.0060	I	0.0787	P	0.0161	W	0.0078
C	0.0282	J	0.0006	Q	0.0007	X	0.0030
D	0.0483	K	0.0064	R	0.0751	Y	0.0168
E	0.1566	L	0.0396	S	0.0715	Z	0.0010
F	0.0167	M	0.0236	T	0.0773		
G	0.0216	N	0.0814	U	0.0272		

Figure 1: The initial probabilities π

Variant 3: Markov Chains

The matrix $P = (p(s|t))_{s,t\in\mathbb{Z}_m}$ describes the transition probabilities; i.e., entry $p(s|t)$ is the *conditional* probability for the event to obtain letter s under the condition that the previously obtained letter was t. Thus, P can also be computed by an appropriate frequency analysis. We may use the sample text exploited to obtain the 2-gram counts. Let $N(i,j)$ denote the number of occurrences of the 2-gram (i,j) in the sample text. We set $N(i) =_{df} \sum_{j=0}^{m-1} N(i,j)$ and compute $p(j|i) = N(i,j)/N(i)$. For example, $N(A,G) = 151$ as displayed in the Figure given above. Moreover, $N(A) = N(A,A) + N(A,B) + \ldots + N(A,Z) = 6476$. Thus, $p(G|A) = 151/6476 = 0.0233$. We leave it to the reader to compute the whole matrix. Now, it is immediately clear that Properties (α), (β), and (γ) are fulfilled.

Variant 3: Markov Chains

We sketch the announced method for computing π. The key property applied here is (δ). In matrix notation, (δ) reads as $\pi = \pi P$. Furthermore, we define $\pi^{(\ell)}(j)$ to be the probability that the Markov chain is in state j at the ℓth step, i.e., $\pi^{(\ell)}(j) = \Pr[X_{\ell-1} = j]$. By definition, $\pi^{(0)}(j) = \pi(j)$.

Assuming (δ), one can prove the remarkable result that $\pi^{(\ell)}(j) = \pi(j)$, too.
That is, the probabilities $\pi^{(\ell)}(j)$, $\ell \in \mathbb{N}$, *do not change* with time, but are stationary. Thus, π can be computed by solving the matrix equation $\pi = \pi P$. It is beyond the scope of this lecture to prove this result here. Instead, the interested reader is encouraged to consult Feller (1968).

Variant 3: Markov Chains

Finally, it is easy to see that the Markov Chain Model is not perfect either. As an easy calculation shows, $\Pr_{\text{plain}}(\text{HELP}) < \Pr_{\text{plain}}(\text{HEPL})$, since $p(P|L) = 0.0041$ and $p(L|P) = 0.0812$. But sometimes it is better than the 2-gram source model. Consider the English word "gaga," and the non-English string "agag." The 2-gram source model gives $\Pr_{\text{plain}}(GAGA) = 0.00000324 < 0.00000484 = \Pr_{\text{plain}}(AGAG)$, while in the Markov Chain Model

$$\begin{aligned}
\Pr_{\text{plain}}(GAGA) &= \pi(G)p(A|G)p(G|A)p(A|G) \\
&= 0.0216 \cdot 0.1078^2 \cdot 0.0233 \\
&= 0.000005848 > 0.000004231 \\
&= \pi(A) \cdot p(G|A) \cdot p(A|G) \cdot p(G|A) \\
&= \Pr_{\text{plain}}(AGAG)\,.
\end{aligned}$$

Variant 3: Markov Chains

The following exercise points to further generalizations of the 1-gram Markov Chain Model:

Exercise 3. *Generalize Definition 5 to the case that X_i depends on $(X_{i-1}, X_{i-2}, \ldots, X_{i-k+1})$, i.e., on the previous $k-1$ letters.*

Thank you!