DD2448 Foundations of Cryptography
Lecture 1

Douglas Wikström
KTH Royal Institute of Technology
dog@kth.se

March 17, 2020

Introduction and Administration

Information About the Course

- **Information given and agreements made during lectures.**
- **Your group members and friends if you miss a lecture.**
- `https://www.kth.se/social/course/DD2448`
- `https://kth.instructure.com/courses/17092`
- Read your KTH email: `<username>@kth.se`
- You cannot use "Discussions" at the course page at Canvas and there is no official course group in any other social media.

If this fails, then email `dog@kth.se` and **use** `DD2448` **in the subject line.**

What is cryptography?

Cryptography is concerned with the conceptualization, definition, and construction of computing systems that address security concerns.

– Oded Goldreich, Foundations of Cryptography, 1997

Applications of Cryptography

Historically.

- Military and diplomatic secret communication.
- Communication between banks, e.g., credit card transactions.

Modern Time.

- Protecting satellite TV from leaching.
- Secrecy and authenticity on the Internet, mobile phones, etc.
- Credit cards.

Applications of Cryptography

Today.

- Distributed file systems, authenticity of blocks in bit torrents, anonymous remailers, Tor-network, etc.
- RFID tags, Internet banking, Försäkringskassan, Skatteverket, "e-legitimation".

Future.

- Secure distributed computing (multiparty computation): election schemes, auctions, secure cloud computing, etc.
- Variations of signatures, cryptosystem, and other primitives with special properties, e.g., group signatures, identity based encryption, etc.

Intended Learning Outcomes

After a completed course, the student should be able to discuss the following basic concepts in cryptography:

- symmetric and asymmetric encryption, digital signatures, cryptographic hash functions and strong pseudorandom generators and to give examples of instantiations of each concept
- conduct simple analyses of cryptographic constructions such as cryptosystems and cryptographic protocols
- read analyses performed by others of cryptographic constructions such as cryptosystems and cryptographic protocols and decide if the given analysis can be trusted
- read and understand technical articles in cryptography.

Recommended Prerequisites

Knowledge equivalent to either one of the courses DD1352 Algorithms, Data Structures and Complexity or DD2354 Algorithms and Complexity and knowledge of probability theory, mathematics and algorithm theory acquired in the mandatory courses of the D or F program.

Tentative Plan of Content (1/2)

- Administration, introduction, classical cryptography.
- Discussion about Group project.
- Symmetric ciphers, substitution-permutation networks, linear cryptanalysis, differential cryptanalysis.
- AES, Feistel networks, DES, modes of operations, DES-variants.
- Entropy and perfect secrecy.
- Security notions of hash functions, random oracles, iterated constructions, SHA, universal hash functions.
- Public-key cryptography, RSA, primality testing, textbook RSA, CPA security.

Tentative Plan of Content (2/2)

- RSA in ROM, Rabin, discrete logarithms, Diffie-Hellman, El Gamal.
- Message authentication codes, identification schemes, signature schemes, PKI.
- Elliptic curve cryptography.
- Pseudorandom generators.
- Post-quantum cryptography
- Cryptographic protocols
- Guest lecture?
- Make-up time and/or special topic.

Course Requirements (1/3)

Group Project. Gives gives G-points. In groups of three students:

- Analyze the pros and cons of using backdoors in cryptographic algorithms and debugging implementations to be able to, e.g., install malware (aka hacking).
- The former is not uncommon historically (albeit rarely openly), and Sweden recently passed a new law to allow the latter: *Hemlig dataavläsning*, Proposition 2019/20:64, which comes into force April 1, 2020.
- Each group must have at least one member who reads Swedish to be able to access Swedish sources.

Detailed rules and advice are found at Canvas.

Course Requirements (2/3)

Homework. Gives T-points.

- Discussed in groups of up to three students.
- Only **informal** discussions are allowed.
- Each student writes and submits their own solution.
- At least one week to complete. We agree on a suitable week in class to avoid conflicts with other courses.

To prepare, a large number of similar or even identical exercises will be shared early in the course, which may be discussed within the same group of three students.

Detailed rules and advice are found at Canvas.

Course Requirements (3/3)

Oral Exam. The purpose is to give a fair grade.

The discussion is based on submitted solutions and aims to verify that the grading corresponds to the learning outcomes of the student.

- G-points or T-points may be added or removed from the tentative grading depending on the understanding shown.
- You can never get a negative number of points on a problem after the oral exam.
- A single O-point is awarded after passing the exam.

Deadlines

The deadlines in this course are strict. Late solutions are not graded.

We negotiate the deadlines to not conflict unnecessarily with other courses.

Solutions are submitted physically during lectures and must adhere the the rules to be graded.

Grading

To earn a given grade the requirements of all lower grades must be satisfied as well.

Grade	O	A/A_0	G/G_0	T/T_0
E	1	0.5	0.4	0.4
D		0.6		
C		0.7	0.5	0.5
B		0.8		0.6
A		0.8		0.8

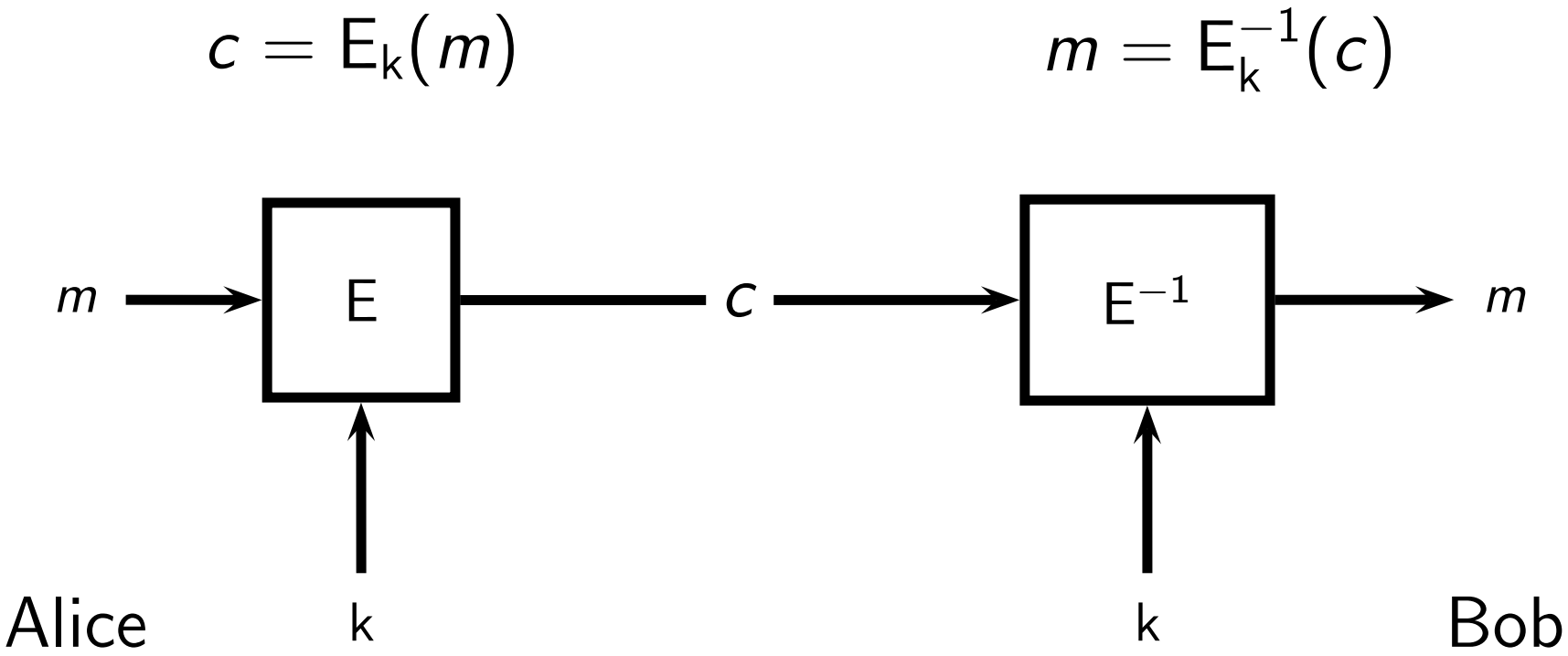
A zero-subscript means nominal number of points, and $A = G + T$. See the course description for details.

Latex

- Latex is the standard typesetting tool for mathematics.
- It is the fastest way to produce mathematical writing.
- **You must use our templates to typeset your solutions.**
- Templates are published at Canvas.
- The best way to learn it is to read:
  `http://tobi.oetiker.ch/lshort/lshort.pdf`

Introduction to Ciphers

Cipher (Symmetric Cryptosystem)



Cipher (Symmetric Cryptosystem)

Definition. A cipher (symmetric cryptosystem) is a tuple $(\mathsf{Gen}, \mathcal{P}, \mathsf{E}, \mathsf{E}^{-1})$, where

Cipher (Symmetric Cryptosystem)

Definition. A cipher (symmetric cryptosystem) is a tuple $(\mathsf{Gen}, \mathcal{P}, \mathsf{E}, \mathsf{E}^{-1})$, where

- Gen is a probabilistic **key generation algorithm** outputting keys from a key space $\mathcal{K}$,

Cipher (Symmetric Cryptosystem)

Definition. A cipher (symmetric cryptosystem) is a tuple $(\mathsf{Gen}, \mathcal{P}, \mathsf{E}, \mathsf{E}^{-1})$, where

- Gen is a probabilistic **key generation algorithm** outputting keys from a key space $\mathcal{K}$,
- $\mathcal{P}$ is a **set of plaintexts**,

Cipher (Symmetric Cryptosystem)

Definition. A cipher (symmetric cryptosystem) is a tuple $(\mathsf{Gen}, \mathcal{P}, \mathsf{E}, \mathsf{E}^{-1})$, where

- Gen is a probabilistic **key generation algorithm** outputting keys from a key space $\mathcal{K}$,
- $\mathcal{P}$ is a **set of plaintexts**,
- E is a deterministic **encryption algorithm**, and

Cipher (Symmetric Cryptosystem)

Definition. A cipher (symmetric cryptosystem) is a tuple $(\mathsf{Gen}, \mathcal{P}, \mathsf{E}, \mathsf{E}^{-1})$, where

- Gen is a probabilistic **key generation algorithm** outputting keys from a key space $\mathcal{K}$,
- $\mathcal{P}$ is a **set of plaintexts**,
- E is a deterministic **encryption algorithm**, and
- E^{-1} is a deterministic **decryption algorithm**,

Cipher (Symmetric Cryptosystem)

Definition. A cipher (symmetric cryptosystem) is a tuple $(\mathsf{Gen}, \mathcal{P}, \mathsf{E}, \mathsf{E}^{-1})$, where

- Gen is a probabilistic **key generation algorithm** outputting keys from a key space $\mathcal{K}$,
- $\mathcal{P}$ is a **set of plaintexts**,
- E is a deterministic **encryption algorithm**, and
- E^{-1} is a deterministic **decryption algorithm**,

such that $\mathsf{E}_\mathsf{k}^{-1}(\mathsf{E}_\mathsf{k}(m)) = m$ for every message $m \in \mathcal{P}$ and $\mathsf{k} \in \mathcal{K}$. The set $\mathcal{C} = \{\mathsf{E}_\mathsf{k}(m) \mid m \in \mathcal{P} \wedge \mathsf{k} \in \mathcal{K}\}$ called the **set of ciphertexts**.

Attacks

Throughout the course we consider various attacks on cryptosystems. With small changes, these attacks make sense both for symmetric and asymmetric cryptosystems.

- Ciphertext-only attack.
- Known-plaintext attack
- Chosen-plaintext attack
- Chosen-ciphertext attack

Ceasar Cipher (Shift Cipher)

Consider English, with alphabet A-Z_, where _ denotes space, thought of as integers 0-26, i.e., $\mathbb{Z}_{27}$

- **Key.** Random letter $k \in \mathbb{Z}_{27}$.
- **Encrypt.** Plaintext $m = (m_1, \ldots, m_n) \in \mathbb{Z}_{27}^n$ gives ciphertext $c = (c_1, \ldots, c_n)$, where $c_i = m_i + k \bmod 27$.
- **Decrypt.** Ciphertext $c = (c_1, \ldots, c_n) \in \mathbb{Z}_{27}^n$ gives plaintext $m = (m_1, \ldots, m_n)$, where $m_i = c_i - k \bmod 27$.

Ceasar Cipher Example

Encoding.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	_
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

Key: $G = 6$

Plaintext.	B	R	I	B	E	_	L	U	L	A	_	T	O	_	B	U	Y	_	J	A	S
Plaintext.	01	17	08	01	04	26	11	20	11	00	26	19	14	26	01	20	24	26	09	00	18
Ciphertext.	07	23	14	07	10	05	17	26	17	06	05	25	20	05	07	26	03	05	15	06	24
Ciphertext.	H	X	O	H	K	F	R	_	R	G	F	Z	U	F	H	_	D	F	P	G	Y

Statistical Attack Against Ceasar (1/3)

Decrypt with all possible keys and see if some English shows up, or more precisely...

Statistical Attack Against Ceasar (2/3)

Written English Letter Frequency Table $F[\cdot]$.

A	0.072	J	0.001	S	0.056
B	0.013	K	0.007	T	0.080
C	0.024	L	0.035	U	0.024
D	0.037	M	0.021	V	0.009
E	**0.112**	N	0.059	W	0.021
F	0.020	O	0.066	X	0.001
G	0.018	P	0.017	Y	0.017
H	0.054	Q	0.001	Z	0.001
I	0.061	R	0.053	_	**0.120**

Note that the same frequencies appear in a ciphertext of written English, but in shifted order!

Statistical Attack Against Ceasar (3/3)

- Check that the plaintext of our ciphertext has similar frequencies as written English.
- Find the key k that maximizes the inner product $T(\mathrm{E}_\mathrm{k}^{-1}(C)) \cdot F$, where $T(s)$ and F denotes the frequency tables of the string s and English.

This usually gives the correct key k.

Affine Cipher

Affine Cipher.

- **Key.** Random pair $\mathrm{k} = (a, b)$, where $a \in \mathbb{Z}_{27}$ is relatively prime to 27, and $b \in \mathbb{Z}_{27}$.
- **Encrypt.** Plaintext $m = (m_1, \ldots, m_n) \in \mathbb{Z}_{27}^n$ gives ciphertext $c = (c_1, \ldots, c_n)$, where $c_i = am_i + b \bmod 27$.
- **Decrypt.** Ciphertext $c = (c_1, \ldots, c_n) \in \mathbb{Z}_{27}^n$ gives plaintext $m = (m_1, \ldots, m_n)$, where $m_i = (c_i - b)a^{-1} \bmod 27$.

Relative primality of a and 27 implies that $(a^{-1} \bmod 27)$ exist.

Substitution Cipher

Ceasar cipher and affine cipher are examples of substitution ciphers.

Substitution Cipher.

- **Key.** Random permutation $\sigma \in S$ of the symbols in the alphabet, for some subset S of all permutations.
- **Encrypt.** Plaintext $m = (m_1, \ldots, m_n) \in \mathbb{Z}_{27}^n$ gives ciphertext $c = (c_1, \ldots, c_n)$, where $c_i = \sigma(m_i)$.
- **Decrypt.** Ciphertext $c = (c_1, \ldots, c_n) \in \mathbb{Z}_{27}^n$ gives plaintext $m = (m_1, \ldots, m_n)$, where $m_i = \sigma^{-1}(c_i)$.

Digrams and Trigrams

- A digram is an ordered pair of symbols.
- A trigram is an ordered triple of symbols.
- It is useful to compute frequency tables for the most frequent digrams and trigrams, and not only the frequencies for individual symbols.

Generic Attack Against Substitution Cipher

1. Compute symbol/digram/trigram frequency tables for the candidate language and the ciphertext.
2. Try to match symbols/digrams/trigrams with similar frequencies.
3. Try to recognize words to confirm your guesses (we would use a dictionary (or Google!) here).
4. Backtrack/repeat until the plaintext can be guessed.

This is hard when several symbols have similar frequencies. A large amount of ciphertext is needed. How can we ensure this?

Vigénère

Vigénère Cipher.

- **Key.** $\mathsf{k} = (\mathsf{k}_0, \ldots, \mathsf{k}_{l-1})$, where $\mathsf{k}_i \in \mathbb{Z}_{27}$ is random.
- **Encrypt.** Plaintext $m = (m_1, \ldots, m_n) \in \mathbb{Z}_{27}^n$ gives ciphertext $c = (c_1, \ldots, c_n)$, where $c_i = m_i + \mathsf{k}_{i \bmod l} \bmod 27$.
- **Decrypt.** Ciphertext $c = (c_1, \ldots, c_n) \in \mathbb{Z}_{27}^n$ gives plaintext $m = (m_1, \ldots, m_n)$, where $m_i = c_i - \mathsf{k}_{i \bmod l} \bmod 27$.

More uniform frequency table :-)

Vigénère

Vigénère Cipher.

- **Key.** $\mathsf{k} = (\mathsf{k}_0, \ldots, \mathsf{k}_{l-1})$, where $\mathsf{k}_i \in \mathbb{Z}_{27}$ is random.
- **Encrypt.** Plaintext $m = (m_1, \ldots, m_n) \in \mathbb{Z}_{27}^n$ gives ciphertext $c = (c_1, \ldots, c_n)$, where $c_i = m_i + \mathsf{k}_{i \bmod l} \bmod 27$.
- **Decrypt.** Ciphertext $c = (c_1, \ldots, c_n) \in \mathbb{Z}_{27}^n$ gives plaintext $m = (m_1, \ldots, m_n)$, where $m_i = c_i - \mathsf{k}_{i \bmod l} \bmod 27$.

More uniform frequency table :-)

We could even make a variant of Vigénère based on the affine cipher, **but is Vigénère really any better than Ceasar?**

Attack Vigénère (1/2)

Index of Coincidence.

- Each probability distribution $p_1, \ldots, p_n$ on n symbols may be viewed as a point $p = (p_1, \ldots, p_n)$ on a $n-1$ dimensional hyperplane in $\mathbb{R}^n$ orthogonal to the vector $\overline{1}$
- Such a point $p = (p_1, \ldots, p_n)$ is at distance $\sqrt{F(p)}$ from the origin, where $F(p) = \sum_{i=1}^{n} p_i^2$.
- It is clear that p is closest to the origin, when p is the uniform distribution, i.e., when $F(p)$ is minimized. (Draw picture!)
- $F(p)$ is invariant under permutation of the underlying symbols $\longrightarrow$ tool to check if a set of symbols is the result of **some** substitution cipher (for non-uniform plaintext sources).

Attack Vigénère (2/2)

1. For $l = 1, 2, 3, \ldots$, we form

$$\begin{pmatrix} C_0 \\ C_1 \\ \vdots \\ C_{l-1} \end{pmatrix} = \begin{pmatrix} c_0 & c_l & c_{2l} & \cdots \\ c_1 & c_{l+1} & c_{2l+1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ c_{l-1} & c_{2l-1} & c_{3l-1} & \cdots \end{pmatrix}$$

   and compute $f_l = \frac{1}{l} \sum_{i=0}^{l-1} F(C_i)$.

2. The local maximum with smallest l is probably the right length.
3. Then attack each C_i separately to recover k_i, using the attack against the Ceasar cipher.

Hill Cipher

Hill Cipher.

- **Key.** k $= A$, where A is an invertible $l \times l$-matrix over $\mathbb{Z}_{27}$.
- **Encrypt.** Plaintext $m = (m_1, \ldots, m_n) \in \mathbb{Z}_{27}^n$ gives ciphertext $c = (c_1, \ldots, c_n)$, where (computed modulo 27):

$$(c_{i+0}, \ldots, c_{i+l-1}) = (m_{i+0}, \ldots, m_{i+l-1})A \ .$$

- **Decrypt.** Ciphertext $c = (c_1, \ldots, c_n) \in \mathbb{Z}_{27}^n$ gives plaintext $m = (m_1, \ldots, m_n)$, where (computed modulo 27):

$$(m_{i+0}, \ldots, m_{i+l-1}) = (c_{i+0}, \ldots, c_{i+l-1})A^{-1} \ .$$

for $i = 1, l+1, 2l+1, \ldots$

Hill Cipher

Hill Cipher.

- **Key.** k $= A$, where A is an invertible $l \times l$-matrix over $\mathbb{Z}_{27}$.
- **Encrypt.** Plaintext $m = (m_1, \ldots, m_n) \in \mathbb{Z}_{27}^n$ gives ciphertext $c = (c_1, \ldots, c_n)$, where (computed modulo 27):

$$(c_{i+0}, \ldots, c_{i+l-1}) = (m_{i+0}, \ldots, m_{i+l-1})A \ .$$

- **Decrypt.** Ciphertext $c = (c_1, \ldots, c_n) \in \mathbb{Z}_{27}^n$ gives plaintext $m = (m_1, \ldots, m_n)$, where (computed modulo 27):

$$(m_{i+0}, \ldots, m_{i+l-1}) = (c_{i+0}, \ldots, c_{i+l-1})A^{-1} \ .$$

for $i = 1, l+1, 2l+1, \ldots$

The Hill cipher is easy to break using a known plaintext attack.

Permutation Cipher (Transposition Cipher)

The permutation cipher is a special case of the Hill cipher.

Permutation Cipher.

- **Key.** Random permutation $\pi \in S$ for some subset S of the set of permutations of $\{0, 1, 2, \ldots, l-1\}$.
- **Encrypt.** Plaintext $m = (m_1, \ldots, m_n) \in \mathbb{Z}_{27}^n$ gives ciphertext $c = (c_1, \ldots, c_n)$, where $c_i = m_{\lfloor i/l \rfloor + \pi(i \bmod l)}$.
- **Decrypt.** Ciphertext $c = (c_1, \ldots, c_n) \in \mathbb{Z}_{27}^n$ gives plaintext $m = (m_1, \ldots, m_n)$, where $m_i = c_{\lfloor i/l \rfloor + \pi^{-1}(i \bmod l)}$.

Summary of Simple Ciphers

- Caesar cipher and affine cipher: $m_i \mapsto am_i + b$.
- Substitution cipher (generalize Ceasar/affine):

$$m_i \mapsto \sigma(m_i)$$

- Vigénère cipher (more uniform frequency table):

$$m_i \mapsto m_i + k_{i \bmod l}$$

- Hill cipher (invertible linear map):

$$(m_1, \ldots, m_l) \mapsto (m_1, \ldots, m_l)A$$

- Transposition cipher (permutation):

$$(m_1, \ldots, m_l) \mapsto (m_{\pi(1)}, \ldots, m_{\pi(l)})$$
$$(m_1, \ldots, m_l) \mapsto (m_1, \ldots, m_l)M_\pi \qquad \text{(equivalently)}$$