Die Geheimſchriften

und

die Dechiffrir-Kunſt.

Mit beſonderer Berückſichtigung
der deutſchen und der franzöſiſchen Sprache

von

F. W. Kaſiski,
Major z. D.

Berlin, 1863.

Druck und Verlag von E. S. Mittler und Sohn.
(Zimmerſtraße 84. 85.)

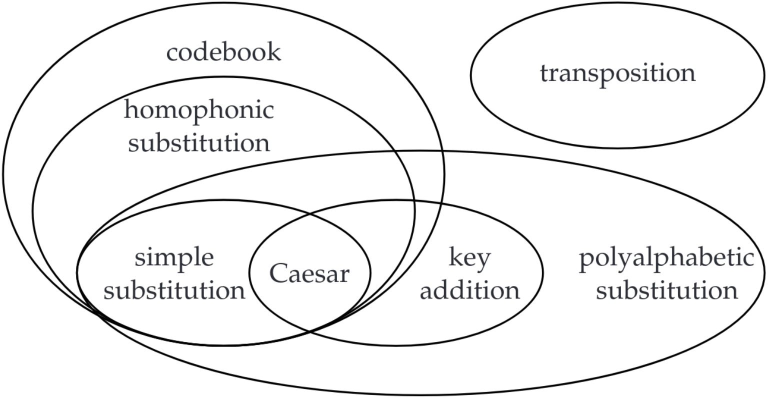
Image Source: Cryptography and the Art of Decryption by Friedrich W. Kasiski

CS 553
CRYPTOGRAPHY

Lecture 3
More Historical Ciphers +
Attack models

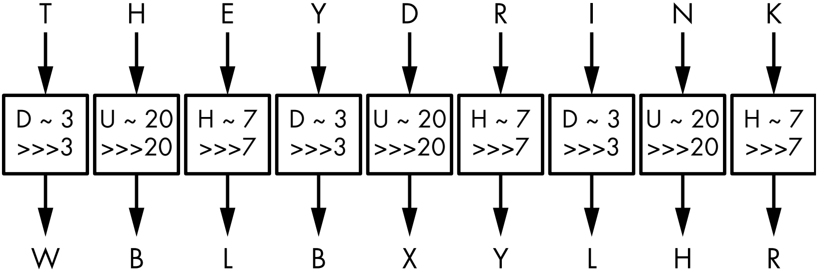
Instructor
Dr. Dhiman Saha

A Taxonomy of Basic Cryptosystems



Vigenere cipher

- By Giovan Battista Bellaso in 16th Century
- Used in American Civil War and World-War I
- Polyalphabetic Substitution



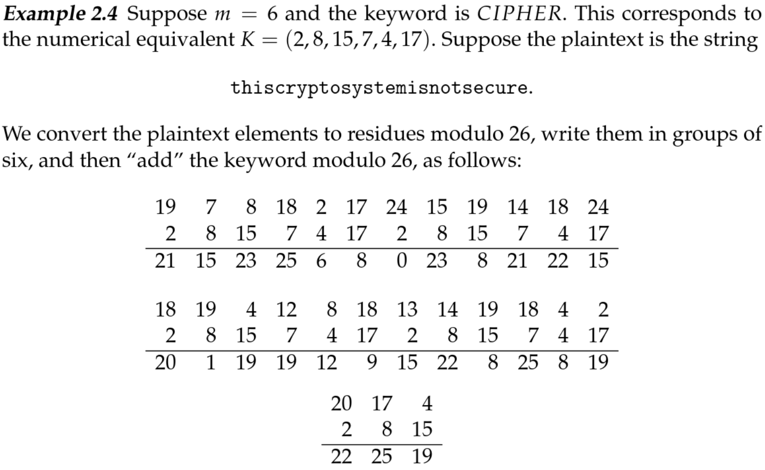
T H E Y D R I N K

D ~ 3 >>>3 | U ~ 20 >>>20 | H ~ 7 >>>7 | D ~ 3 >>>3 | U ~ 20 >>>20 | H ~ 7 >>>7 | D ~ 3 >>>3 | U ~ 20 >>>20 | H ~ 7 >>>7

W B L B X Y L H R

Key = ‘DUH’

Another Example



Example 2.4 Suppose $m = 6$ and the keyword is *CIPHER*. This corresponds to the numerical equivalent $K = (2, 8, 15, 7, 4, 17)$. Suppose the plaintext is the string

`thiscryptosystemisnotsecure.`

We convert the plaintext elements to residues modulo 26, write them in groups of six, and then "add" the keyword modulo 26, as follows:

$$\begin{array}{rrrrrrrrrrrr}
19 & 7 & 8 & 18 & 2 & 17 & 24 & 15 & 19 & 14 & 18 & 24 \\
2 & 8 & 15 & 7 & 4 & 17 & 2 & 8 & 15 & 7 & 4 & 17 \\
\hline
21 & 15 & 23 & 25 & 6 & 8 & 0 & 23 & 8 & 21 & 22 & 15
\end{array}$$

$$\begin{array}{rrrrrrrrrrrr}
18 & 19 & 4 & 12 & 8 & 18 & 13 & 14 & 19 & 18 & 4 & 2 \\
2 & 8 & 15 & 7 & 4 & 17 & 2 & 8 & 15 & 7 & 4 & 17 \\
\hline
20 & 1 & 19 & 19 & 12 & 9 & 15 & 22 & 8 & 25 & 8 & 19
\end{array}$$

$$\begin{array}{rrr}
20 & 17 & 4 \\
2 & 8 & 15 \\
\hline
22 & 25 & 19
\end{array}$$

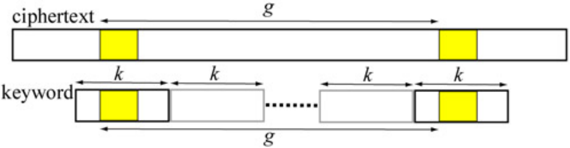
- Ciphertext → `VPXZGIAXIVWPUBTTMJPWIZITWZT`
- What is the size of the key-space?

Kasiski Test

How to break it?

Intuition

If a repeated substring in a plaintext is encrypted by the same substring in the keyword, then the ciphertext contains a repeated substring and the distance of the two occurences is a multiple of the keyword length.



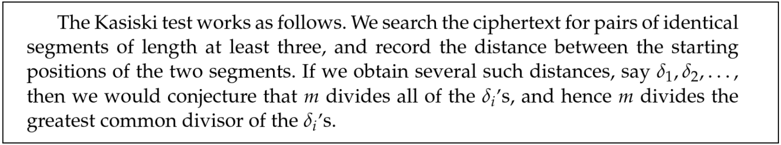
THEY DRINK THE TEA → WBLBXYLHRWBLWYH

- Note: Since the keyword of length k is repeated to fill the length of the ciphertext, the distance g is a multiple of the keyword length k.
- For example, if the distance is $g = 18$, since the factors of g are $2, 3, 6, 9$ and 18, one of them may be the length of the unknown keyword.

Image Source: https://pages.mtu.edu/~shene/NSF-4/Tutorial/VIG/Vig-Kasiski.html

Kasiski Test

- Friedrich W. Kasiski, a German military officer, published his book *Die Geheimschriften und die Dechiffrirkunst* (Cryptography and the Art of Decryption) in 1863



The Kasiski test works as follows. We search the ciphertext for pairs of identical segments of length at least three, and record the distance between the starting positions of the two segments. If we obtain several such distances, say $\delta_1, \delta_2, \ldots,$ then we would conjecture that m divides all of the δ_i's, and hence m divides the greatest common divisor of the δ_i's.

- Still difficult to break for short messages
- Good for short-lived messages

Trivia

The 19th-century cryptographer Auguste Kerckhoffs estimated that most encrypted wartime messages required confidentiality for only three to four hours

Hill Cipher

- Polyalphabetic; by Lester S. Hill. in 1929
- $\mathcal{P} = \mathcal{C} = (\mathbb{Z}_{26})^m$
- Let, $(x_1, x_2, \cdots, x_m) \in \mathcal{P}$, $K = (k_{i,j}) \in \mathcal{K}$ then
  $y = e_K(x) = (y_1, y_2, \cdots, y_m)$

$$(y_1, y_2, \cdots, y_m) = (x_1, x_2, \cdots, x_m) \begin{bmatrix} k_{1,1} & k_{1,2} & \cdots & k_{1,m} \\ k_{2,1} & k_{2,2} & \cdots & k_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ k_{m,1} & k_{m,2} & \cdots & k_{m,m} \end{bmatrix}$$

- Note: every y_i is a linear combination of all x_i ⚠
- In matrix notation:
  Encryption $y = xK$ and Decryption $x = yK^{-1}$
- Invertibility of K

Permutation/Transposition Cipher

Definition

Let m be a positive integer. Let $\mathcal{P} = \mathcal{C} = (\mathbb{Z}_{26})^m$ and let $\mathcal{K}$ consist of all permutations of $1, \cdots, m$. For a key π, we define

$$e_\pi(x_1, x_2, \cdots, x_m) = x_{\pi(1)}, x_{\pi(2)}, \cdots, x_{\pi(m)}$$

and

$$d_\pi(y_1, y_2, \cdots, y_m) = x_{\pi^{-1}(1)}, x_{\pi^{-1}(2)}, \cdots, x_{\pi^{-1}(m)}$$

where π^{-1} is the inverse permutation of π.

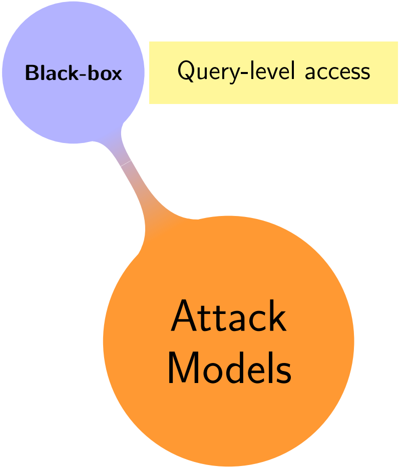
Example

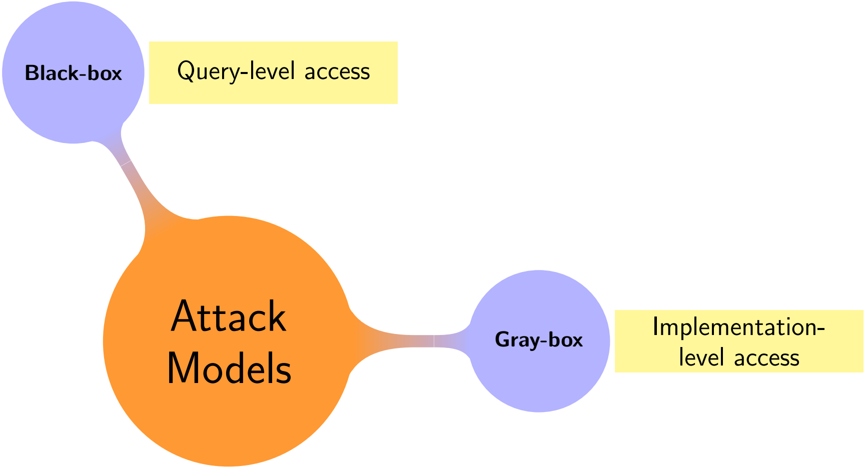
Let $m = 6$ and a possible key is the permutation π:

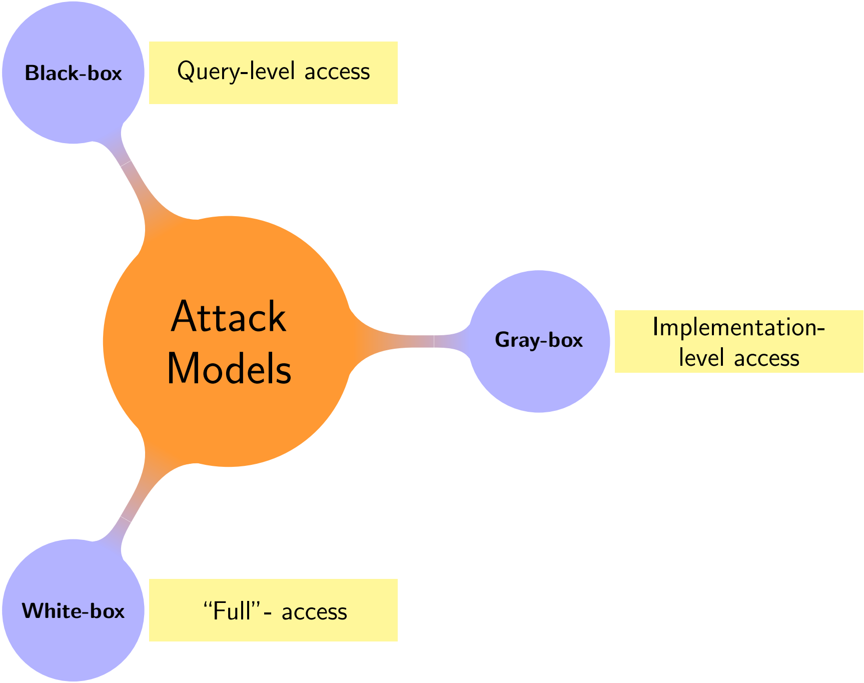
x	1	2	3	4	5	6
$\pi(x)$	3	5	1	6	4	2

- This is a special case of Hill Cipher. How?

Attack
Models


Black-box
Query-level access
Attack Models


Attack Models
Black-box
Query-level access
Gray-box
Implementation-level access

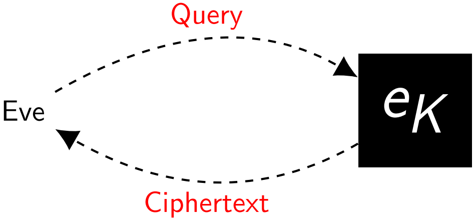

Black-box
Query-level access
Attack Models
Gray-box
Implementation-level access
White-box
“Full”- access

Black-box Models

Notion of query — In current context

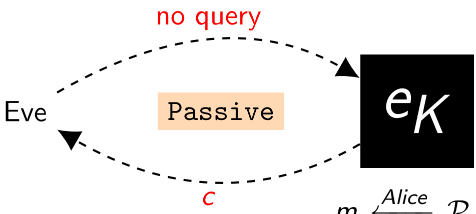
Operation that sends an input value to some function and gets the output in return, without exposing the details of that function.

- Assume cipher behaving as a black-box
- Eve only sees what goes in and out of the cipher
- Different models based on what Eve can query for



Ciphertext Only Attack (COA)

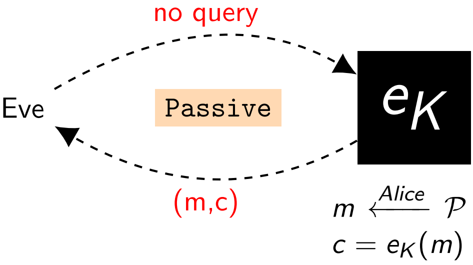
- Eve has access to the ciphertext c only
- But **does not** know the associated plaintexts
- Or how they were selected
- The default scenario
- Weakest adversarial model



$$m \xleftarrow{Alice} \mathcal{P}$$
$$c = e_K(m)$$

Known Plaintext Attack (KPA)

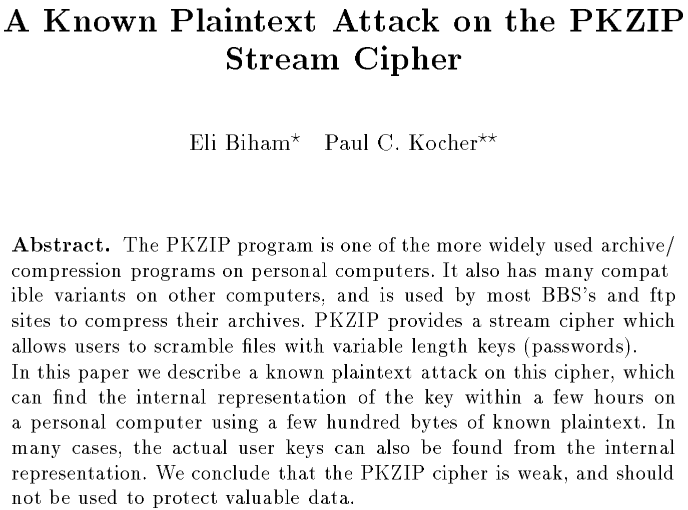
- Eve can get her hands on both a ciphertext c and its corresponding plaintext m,
- Where plaintexts are **assumed to be randomly selected.**
- Stronger than COA
- Is Hill Cipher KPA-vulnerable? △



$m \xleftarrow{Alice} \mathcal{P}$

$c = e_K(m)$

Practical Example



A Known Plaintext Attack on the PKZIP Stream Cipher

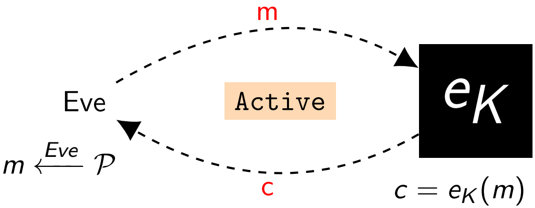
Eli Biham* Paul C. Kocher**

Abstract. The PKZIP program is one of the more widely used archive/compression programs on personal computers. It also has many compatible variants on other computers, and is used by most BBS's and ftp sites to compress their archives. PKZIP provides a stream cipher which allows users to scramble files with variable length keys (passwords).
In this paper we describe a known plaintext attack on this cipher, which can find the internal representation of the key within a few hours on a personal computer using a few hundred bytes of known plaintext. In many cases, the actual user keys can also be found from the internal representation. We conclude that the PKZIP cipher is weak, and should not be used to protect valuable data.

`https://www.unix-ag.uni-kl.de/~conrad/krypto/pkcrack.html`

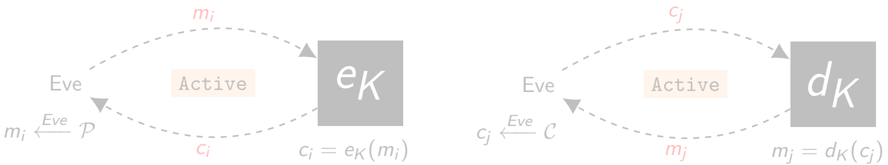
Chosen Plaintext Attack (CPA)

- Eve can perform encryption queries for plaintexts of her choice
- And observe the resulting ciphertexts.
- Notion of **active** adversary
- Stronger than KPA
- Affine Cipher is CPA-vulnerable. Why?



Chosen Ciphertext Attack (CCA)

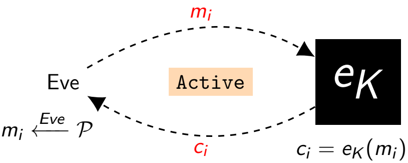
- Eve can perform encryption as well as decryption queries.
- Does this make sense?
- Decrypting something is not always enough to break a system.
- Stronger than CPA

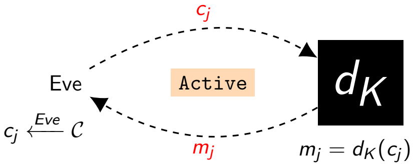


- More advanced: Adaptive CCA

Chosen Ciphertext Attack (CCA)

- Eve can perform encryption as well as decryption queries.
- Does this make sense?
- Decrypting something is not always enough to break a system.
- Stronger than CPA





- More advanced: Adaptive CCA

Formally

KPA, CPA, CCA

KPA-Security

Given one or more pairs of plaintexts and their corresponding ciphertexts, $(m_1, c_1), (m_2, c_2), \cdots, (m_n, c_n)$ it must be very difficult to decrypt any ciphertext c that is **not in the given list** without knowing k.

CPA-Security

For any list of plaintexts $m_1, m_2, \cdots, m_n \in \mathcal{P}$ chosen by the adversary, even with knowledge of the corresponding ciphertexts $e_K(m_1), e_K(m_2), \cdots, e_K(m_n)$ it is very difficult to decrypt any ciphertext c that is **not in the given list** without knowing k.

CCA-Security

HW-Problem

Formally — KPA, CPA, CCA

KPA-Security

Given one or more pairs of plaintexts and their corresponding ciphertexts, $(m_1, c_1), (m_2, c_2), \cdots, (m_n, c_n)$ it must be very difficult to decrypt any ciphertext c that is **not in the given list** without knowing k.

CPA-Security

For any list of plaintexts $m_1, m_2, \cdots, m_n \in \mathcal{P}$ chosen by the adversary, even with knowledge of the corresponding ciphertexts $e_K(m_1), e_K(m_2), \cdots, e_K(m_n)$ it is very difficult to decrypt any ciphertext c that is **not in the given list** without knowing k.

CCA-Security

HW-Problem

Formally

KPA-Security

Given one or more pairs of plaintexts and their corresponding ciphertexts, $(m_1, c_1), (m_2, c_2), \cdots, (m_n, c_n)$ it must be very difficult to decrypt any ciphertext c that is **not in the given list** without knowing k.

CPA-Security

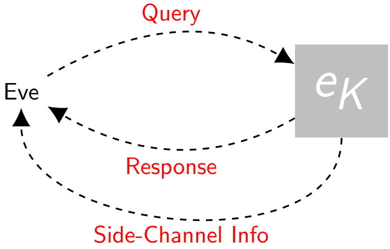
For any list of plaintexts $m_1, m_2, \cdots, m_n \in \mathcal{P}$ chosen by the adversary, even with knowledge of the corresponding ciphertexts $e_K(m_1), e_K(m_2), \cdots, e_K(m_n)$ it is very difficult to decrypt any ciphertext c that is **not in the given list** without knowing k.

CCA-Security

HW-Problem △

Gray-box Models

- Eve has access to a ciphers **implementation**.
- Captures practical scenarios
- Models physical access for applications such as smart cards, embedded systems
- **Notion of Side-channel attacks**[1]
- Software - timing, errors, return values
- Hardware - power, EM, faults
- Non-invasive/Invasive



[1] Will be explored later in the course

White-Box Models

- Worst-case attack model
- Eve has **full control** over the cryptographic program and its **execution environment**
- Consider ability to decrypt ciphertexts under a certain key **without sharing the key itself** - Applications like DRM
- How is that even possible?

How about embedding the key in the code itself?

Are you serious!!!

- What if this is reverse engineered?

White-Box Models

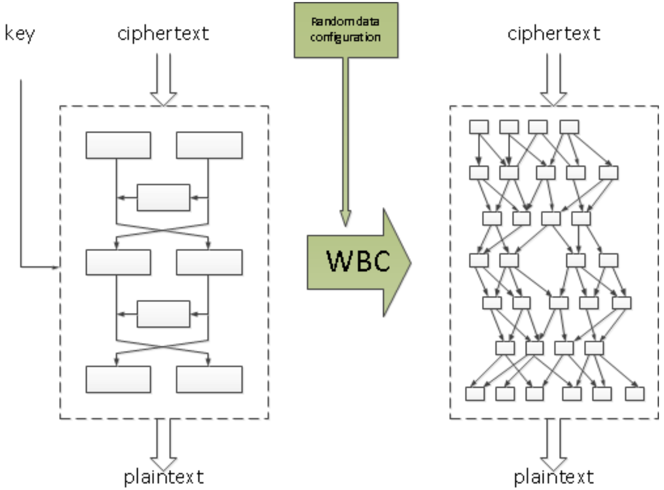
- Worst-case attack model
- Eve has **full control** over the cryptographic program and its **execution environment**
- Consider ability to decrypt ciphertexts under a certain key **without sharing the key itself** - Applications like DRM
- How is that even possible?

How about embedding the key in the code itself?

Are you serious!!!

- What if this is reverse engineered?

White-Box Crypto



- Idea: obfuscate the code to prohibit reverse engineering

Image Source: http://www.whiteboxcrypto.com/