

# CPSC 418/MATH 318 Introduction to Cryptography

## Classical Ciphers, Perfect Secrecy, One-Time Pad

Renate Scheidler

Department of Mathematics & Statistics  
Department of Computer Science  
University of Calgary

Week 2

“Few false ideas have more firmly gripped the minds of so many intelligent men than the one that, if they just tried, they could invent a cipher that no one could break.”

David Kahn, *The Code Breakers*, 1967

## Outline

- 1 Cryptanalysis
  - Cryptographic Attacks
  - Cryptographic Security
- 2 Terminology of Security
- 3 Historical Ciphers
- 4 Probability Theory
- 5 Perfect Secrecy
- 6 Vernam One-Time Pad

## Goals of an Attacker

- Deduce the key or portions thereof
- Deduce one or more plaintexts or portions thereof
- Modify a message
- Replay a message
- Impersonate (*i.e.* masquerade as) another entity

The first two are passive attacks, the last three active attacks.

## Attacks – Means of an Attacker(Recap)

- Ciphertext Only Attack (COA)
- Known Plaintext Attack (KPA)
- Chosen Plaintext Attack (CPA)
  - Has an adaptive variant
- Chosen Ciphertext Attack (CCA)
  - Non-adaptive version (CCA1)
  - Adaptive version (CCA2)

The first two are passive attacks, the last three active attacks.

## Notions of Security

### Definition 1 (Kerckhoff's Principle)

The security of a cryptosystem should depend entirely upon knowledge of the key, not of the method.

- From “La Cryptographie Militaire” (1883), one of the first scientific treatments of cryptography.
- This implies in particular that a cipher should be completely published and still be secure (against its own designer and everyone else).
- Se skcd cartoon on last week's title slide

So what constitutes a *secure* cryptosystem? We saw that a good system should be secure against adaptive CCA's. What does “secure” mean? There are different notions of security.

## Notions of Security

Listed from strongest to weakest:

- *Unconditional* security – can an adversary with unlimited computing power defeat the system?
- *Provable* security – breaking the system can be reduced (mathematically) to another, supposedly difficult problem; e.g. integer factorization.
- *Computational* security – does the perceived amount of computing power necessary to break the system (using the best known method) exceed (by a comfortable margin) the available computing power of the attacker?
- *Ad-hoc* security – security is argued via a series of convincing arguments that every successful attack is impractical.
  - Entirely unacceptable in professional crypto

## Remarks

Computational security often used in conjunction with provable security

- E.g. a typical security claim might read something like “a cryptosystem is provably secure against an adaptive CCA, in the standard model, assuming integer factorization is intractable”

Provable security does *not* mean that a cryptosystem is *proved* secure!

- Proofs typically only reduce to another problem (which could eventually be solved)
- Proofs assume specific adversarial capabilities and attacks (eg. adaptive CCA). This is called a *proof model*.

## Classical Ciphers

Classical ciphers usually belong to one of the following two types: substitution or transposition ciphers.

### Definition 2 (Substitution cipher)

A cipher for which encryption replaces each plaintext symbol by some ciphertext symbol without changing the order of the plaintext symbols.

### Definition 3 (Transposition cipher)

A cipher in which the ciphertext is a rearrangement (*i.e.* permutation) of the plaintext symbols.

## Examples of Classical Ciphers

Examples of substitution ciphers:

- Shift cipher: to encrypt, every plaintext letter is shifted by a fixed position  
*monoalphabetic*: one cipher alphabet
- Vigenère cipher: plaintext letters are shifted by different positions based on a repeated rotating pattern (see handouts)  
*polyalphabetic*: several cipher alphabets

Examples of transposition ciphers:

- Route cipher: plaintext is arranged in some geometric figure and encrypted by rearranging the plaintext according to some route through the figure  
e.g. in a *columnar transposition* cipher, the plaintext is arranged in a rectangle and the ciphertext consists of a secret permutation of the plaintext columns

## Past Uses of Substitution Ciphers

History:

- Mary Queen of Scots conspiring to overthrow Queen Elizabeth I and gain the English throne
- Famous 1917 WW I Zimmerman telegram
- *Enigma* machines and Navajo Code talkers in WW II

Literature:

- Edgar Allan Poe's *The Gold Bug*
- Arthur Conan Doyle's *The Adventure of the Dancing Men* (a Sherlock Holmes story)
- Kabbalistic texts, writings of Jewish mysticism and the biblical book of Jeremiah use the *atbash* cipher (encrypts via alphabet reversal)

A pathological example (which would not work for frequency analysis):

- *Gadsby* by Ernest Vincent Wright (1939) is a 50,000 word novel written entirely without using the letter E

## Cryptanalysis of Monoalphabetic Substitution Ciphers

- 1 Highly vulnerable to KPA's: each portion of corresponding plaintext and ciphertext reveals some of the cipher.
  - Eg. For shift ciphers, one letter pair reveals the key!
- 2 Each plaintext letter is encrypted to the same ciphertext letter.
  - Frequent ciphertext letters correspond to common plaintext letters
  - Pairs of identical ciphertext letters correspond to such plaintext letter pairs (e.g. "XX" corresponds to "yy")
- 3 Language redundancy generally yields the key, given a sufficient amount of ciphertext (COA).
  - frequency distribution of the plaintext alphabet (letters, pairs of letters, triples of letters etc.) in a given language can be established statistically and compared with the ciphertext (see frequency and digraph handouts).

## Cryptanalysis of Other Classical Ciphers

Polyalphabetic substitution ciphers and transposition ciphers are also vulnerable to KPAs and COAs.

Cryptanalysis of Vigenère cipher:

- Determine the length of rotation patterns (i.e. the number of cipher alphabets) via guessing, the *kappa* test or *Kasiski's factoring method*
- Cryptanalyze each subtext as a shift cipher

Cryptanalysis of columnar transposition:

- Guess the dimensions of the rectangle
- Determine the order of the columns via frequency counts (which will be the same as for English text). Place columns adjacent to each other if they produce common letter pairs (e.g. QX is extremely unlikely, but EN is highly likely).

## Modern Usage

Individually, substitution ciphers and transposition ciphers are generally insecure.

However, when alternating them repeatedly,

$$M \rightarrow \boxed{T} \rightarrow \boxed{S} \rightarrow \boxed{T} \rightarrow \boxed{S} \rightarrow \dots \rightarrow \boxed{T} \rightarrow \boxed{S} \rightarrow C,$$

they become very secure.

This idea, due to *Claude Shannon*, is the basis of the design of modern symmetric cryptosystems.

## Information Theory

*Claude Shannon* is widely hailed as the “father of information theory”.

- seminal work in the late 1940's and early 1950's in this field
- credited with turning cryptography into a scientific discipline.
- in addition, modern satellite transmission would not be possible without his work

*Information theory* measures the amount of information conveyed by a piece of data.

- captures how much partial information you need to have in order to obtain full information.

## Partial Information

For example, partial information reveals the full word or phrase in:

- Abbreviations — “LOL”
- Contractions — “I've”
- Omitted vowels — “BSKTBLL”
- Glyphs (e.g. emojis) — smiley face

How much partial information is enough? E.g. “BLL” could mean “ball”, “bell”, “bill”, “bull”, ...

## Definitions for Probability Theory

### Definition 4

*Sample space* – a finite set  $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$  whose elements are called *outcomes*.

*Probability distribution* on  $\mathcal{X}$  – a complete set of probabilities; i.e.

$$p(X_1), p(X_2), \dots, p(X_n) \geq 0 \quad \text{with} \quad \sum_{i=1}^n p(X_i) = 1.$$

*Random variable* – a pair  $X$  consisting of a sample space  $\mathcal{X}$  and a probability distribution  $p$  on  $\mathcal{X}$ . The *probability* that  $X$  takes on the value  $x \in \mathcal{X}$  is denoted by  $p(X = x)$  or simply  $p(x)$ .

$p(x)$  is AKA the *a priori* probability of  $x$  (“a priori” = from before)

## Joint and Conditional Probability

Let  $X$  and  $Y$  be random variables,  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ .

### Definition 5

*Joint probability*  $p(x, y)$ : probability that  $p(X = x)$  and  $p(Y = y)$ .

*Conditional probability*  $p(x|y)$ : probability that  $p(X = x)$  given that  $p(Y = y)$ .

$p(x|y)$  is AKA known as the *a posteriori* probability of  $x$  given  $y$  (“a posteriori” = after the fact)

Joint and conditional probabilities are related as follows:

$$p(x, y) = p(x|y)p(y) .$$

## Bayes' Theorem

### Theorem 1 (Bayes Theorem)

If  $p(y) > 0$ , then

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)} .$$

### Proof.

Clearly  $p(x, y) = p(y, x)$ , so  $p(x|y)p(y) = p(y|x)p(x)$ . Now divide by  $p(y)$ . □

## Independence

### Definition 6

Two random variables  $X, Y$  are *independent* if  $p(x, y) = p(x)p(y)$  for all  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ .

### Example 7

A fair coin toss is modeled by a random variable on the sample space  $\mathcal{X} = \{\text{heads}, \text{tails}\}$  so that  $p(\text{heads}) = p(\text{tails}) = 1/2$ . Two fair coin tosses in a row represent independent events as each of the 4 possible outcomes has (joint) probability  $1/4$ .

### Corollary 2

$X$  and  $Y$  are independent if and only if  $p(x|y) = p(x)$  for all  $x \in \mathcal{X}, y \in \mathcal{Y}$  with  $p(y) > 0$ .

## Idea of Perfect Secrecy

Recall the notion of *unconditional security* which requires that an adversary with unlimited computing power cannot defeat the system. This relates to *perfect secrecy*.

Intuitively, for perfect secrecy, ciphertexts should reveal no information whatsoever about plaintexts.

Theoretically unbreakable!

## Setup

We consider the following three probability distributions:

- A random variable on the message space  $\mathcal{M}$ ; plaintexts  $M$  occur with probabilities  $p(M)$  such that  $\sum_{M \in \mathcal{M}} p(M) = 1$ .
- A random variable on the ciphertext space  $\mathcal{C}$ ; ciphertexts  $C$  occur with probabilities  $p(C)$  such that  $\sum_{C \in \mathcal{C}} p(C) = 1$ .
- A random variable on the key space  $\mathcal{K}$ ; keys  $K$  are selected with *prior* probabilities  $p(K)$  such that  $\sum_{K \in \mathcal{K}} p(K) = 1$ .

We assume that the random variables on  $\mathcal{K}$  and  $\mathcal{M}$  are independent, as keys are usually chosen before the plaintext is ever seen.

- Most of the time, each key is selected with equal likelihood  $1/|\mathcal{K}|$ , regardless of the nature of the messages to be encrypted.

## Notation

We consider the following probabilities:

- $p(M)$  — probability that plaintext  $M$  occurs
- $p(C)$  — probability that ciphertext  $C$  occurs (as some encryption)
- $p(M|C)$  — probability that  $M$  is the decryption of a given  $C$   
(More formally, that  $M$  is a possible plaintext, given that ciphertext  $C$  is encountered, e.g. received in a transmission)
- $p(C|M)$  — probability that  $C$  is the encryption of a given  $M$   
(More formally, that ciphertext  $C$  was encountered, given that plaintext  $M$  occurs as a possible plaintext)
- $p(K)$  — probability that key  $K$  was chosen

## Definition

### Definition 8 (Perfect Secrecy)

A cryptosystem provides *perfect secrecy* if  $p(M|C) = p(M)$  for all  $M \in \mathcal{M}$  and  $C \in \mathcal{C}$  with  $p(C) > 0$ .

Formally, perfect secrecy means exactly that the random variables on  $\mathcal{M}$  and  $\mathcal{C}$  are independent. Informally, this implies that knowing the ciphertext  $C$  gives us no information about  $M$ .

The probabilities  $p(M|C)$  and  $p(M)$  are hard to quantify (we may not know anything about which plaintexts occur). Bayes' Theorem relates these quantities to  $p(C|M)$  and  $p(C)$ , and these probabilities turn out to be easier to quantify.

## Equivalent Characterization

### Theorem 3

A cryptosystem provides perfect secrecy if and only if  $p(C|M) = p(C)$  for all  $M \in \mathcal{M}, C \in \mathcal{C}$  with  $p(M) > 0$  and  $p(C) > 0$ .

### Proof

By Bayes' Theorem,

$$p(C|M) = \frac{p(C)p(M|C)}{p(M)} \quad (*)$$

for all  $M \in \mathcal{M}, C \in \mathcal{C}$  with  $p(M) > 0, p(C) > 0$ .

" $\Rightarrow$ ": Assume perfect secrecy, and let  $M \in \mathcal{M}, C \in \mathcal{C}$  with  $p(M) > 0$  and  $p(C) > 0$ . Since  $p(M|C) = p(M)$  by perfect secrecy,  $(*)$  yields  $p(C|M) = p(C)$ .

## Proof of Theorem 3 (cont'd)

## Proof (cont'd)

" $\Leftarrow$ ": Assume  $p(C|M) = p(C)$  for all  $M \in \mathcal{M}$ ,  $C \in \mathcal{C}$  with  $p(M) > 0$  and  $p(C) > 0$ . By definition of perfect secrecy, we need to prove that  $p(M|C) = p(M)$  for all  $M \in \mathcal{M}$  and  $C \in \mathcal{C}$  with  $p(C) > 0$ .<sup>a</sup>

So let  $M \in \mathcal{M}$  and  $C \in \mathcal{C}$  with  $p(C) > 0$ .

**Case**  $p(M) > 0$ . Since  $p(C|M) = p(C)$  by assumption,  $(*)$  yields  $p(M|C) = p(M)$  in this case.

**Case**  $p(M) = 0$ . Then<sup>b</sup>  $p(M|C) = 0$ , as the additional restriction that  $C$  is given does not increase the probability. Hence  $p(M|C) = 0 = p(M)$ .  $\square$

<sup>a</sup>This needs to be proved for ALL messages  $M$ , i.e. those with  $p(M) > 0$  and those for which  $p(M) = 0$ .

<sup>b</sup>In this case, the assertion of the theorem is not applicable because it is a statement about messages  $M$  with  $p(M) > 0$  and says nothing about messages  $M$  with  $p(M) = 0$ . So we must prove perfect secrecy by other means.

## Intuition

**Informal interpretation of Theorem 3:** Perfect secrecy means that the probability that a ciphertext  $C$  is the encryption of a particular plaintext  $M$  (under some key  $K$ ) is the same as the probability that  $C$  is the encryption of any other plaintext (possibly enciphered under another key).

In other words,  $M$  is not more likely as a candidate for the decryption of  $C$  than any other plaintext.

Completely foils COAs:  $C$  tells you nothing about  $M$ .

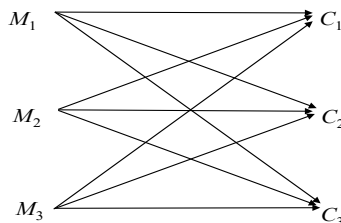
## A Simple Example

Suppose we have 3 messages, i.e.  $\mathcal{M} = \{M_1, M_2, M_3\}$ , and 3 ciphertexts  $\mathcal{C} = \{C_1, C_2, C_3\}$ , and all occur with equal probabilities:

$$p(M_1) = p(M_2) = p(M_3) = 1/3, \quad p(C_1) = p(C_2) = p(C_3) = 1/3.$$

Also, suppose that we have perfect secrecy, i.e.  $p(M_i|C_j) = p(M_i) = 1/3$  for all  $i, j$ . By Theorem 3, we have  $p(C_i|M_j) = p(C_i) = 1/3$  for all  $i, j$ .

This means that each ciphertext  $C_i$  could be the encryption of any of the messages  $M_j$  with equal probability (each arrow is equally likely):

Computing  $p(C|M)$ 

Recall that perfect secrecy is equivalent to  $p(C|M) = p(C)$  for all messages  $M$  and all ciphertexts  $C$  that occur.

How can we determine  $p(C|M)$  and  $p(C)$ ?

For any message  $M \in \mathcal{M}$ , we have

$$p(C|M) = \sum_{\substack{K \in \mathcal{K} \text{ with} \\ E_K(M)=C}} p(K).$$

That is,  $p(C|M)$  is the sum of probabilities  $p(K)$  over all those keys  $K \in \mathcal{K}$  that encipher  $M$  to  $C$ .

## Number of Keys in the Sum

Usually there is at most one key  $K$  with  $E_K(M) = C$  for given  $M$  and  $C$ .

However, some ciphers can transform the same plaintext into the same ciphertext with different keys.

- A monoalphabetic substitution cipher will transform a message into the same ciphertext with different keys if the only differences between the keys occur for characters which do not appear in the message
- Eg. key1 = ECONOMICS, key2 = ECONOMY, and we encrypt a message of at most 6 characters).

## Example: Computing $p(C|M)$

$\mathcal{M} = \{a, b\}$ ,  $\mathcal{K} = \{K_1, K_2, K_3\}$ , and  $\mathcal{C} = \{1, 2, 3, 4\}$ . Encryption is given by the following table:

Key	$M = a$	$M = b$
$K_1$	$C = 1$	$C = 2$
$K_2$	$C = 2$	$C = 3$
$K_3$	$C = 3$	$C = 4$

Thus,

$$\begin{aligned} p(1|a) &= p(K_1), & p(1|b) &= 0, \\ p(2|a) &= p(K_2), & p(2|b) &= p(K_1), \\ p(3|a) &= p(K_3), & p(3|b) &= p(K_2), \\ p(4|a) &= 0, & p(4|b) &= p(K_3). \end{aligned}$$

## Description of $E_K$

Consider a fixed key  $K$ . The mathematical description of the set of all possible encryptions (of any plaintext) under this key  $K$  is exactly the image of  $E_K$ , i.e. the set  $E_K(\mathcal{M}) = \{E_K(M) \mid M \in \mathcal{M}\}$ .

Key	$M = a$	$M = b$
$K_1$	$C = 1$	$C = 2$
$K_2$	$C = 2$	$C = 3$
$K_3$	$C = 3$	$C = 4$

In the previous example, we have

- $E_{K_1}(\mathcal{M}) = \{1, 2\}$
- $E_{K_2}(\mathcal{M}) = \{2, 3\}$
- $E_{K_3}(\mathcal{M}) = \{3, 4\}$ .

## Computation of $p(C)$

For a key  $K$  and ciphertext  $C \in E_K(\mathcal{M})$ , consider the probability  $p(D_K(C))$  that the message  $M = D_K(C)$  was sent. Then

$$p(M|K) = p(M)$$

as the random variables on plaintexts and keys are assumed to be independent.

$$p(C) = \sum_{\substack{K \in \mathcal{K} \text{ with} \\ C \in E_K(\mathcal{M})}} p(D_K(C)|K)p(K) = \sum_{\substack{K \in \mathcal{K} \text{ with} \\ C \in E_K(\mathcal{M})}} p(D_K(C))p(K).$$

That is,  $p(C)$  is the sum of probabilities over all those keys  $K \in \mathcal{K}$  under which  $C$  has a decryption under key  $K$ , each weighted by the probability that that key  $K$  was chosen.

## Example, cont.

Key	$M = a$	$M = b$
$K_1$	$C = 1$	$C = 2$
$K_2$	$C = 2$	$C = 3$
$K_3$	$C = 3$	$C = 4$

The respective probabilities of the four ciphertexts 1, 2, 3, 4 are:

$$p(1) = p(a)p(K_1), \quad p(2) = p(b)p(K_1) + p(a)p(K_2)$$

$$p(3) = p(b)p(K_2) + p(a)p(K_3), \quad p(4) = p(b)p(K_3)$$

If we assume that every key and every message is equally probable, i.e.  $p(K_1) = p(K_2) = p(K_3) = 1/3$  and  $p(a) = p(b) = 1/2$ , then

$$p(1) = (1/2)(1/3) = 1/6, \quad p(2) = 2(1/2)(1/3) = 1/3$$

$$p(3) = 2(1/2)(1/3) = 1/3, \quad p(4) = (1/2)(1/3) = 1/6$$

Note that  $p(1|a) = p(K_1) = 1/3 \neq 1/6 = p(1)$ , so this system does not provide perfect secrecy.

## Necessary Condition for Perfect Secrecy

## Theorem 4

*If a cryptosystem has perfect secrecy, then  $|\mathcal{K}| \geq |\mathcal{M}|$ .*

Informal argument via contradiction: suppose  $|\mathcal{K}| < |\mathcal{M}|$ .

- Pick a ciphertext  $C$  with  $p(C) > 0$  (i.e.  $C$  actually occurs as the encryption of some message under some key).
- Since  $|\mathcal{K}| < |\mathcal{M}|$ , there is some message  $M$  such that no key  $K$  encrypts  $M$  to  $C$  (i.e.  $M$  is ruled out as a possible decryption of  $C$ ).
- This means that the sum defining  $p(C|M)$  is empty, so  $p(C|M) = 0$ .
- Since  $p(C) > 0$ , we have no perfect secrecy.

After intercepting a particular ciphertext  $C$  (i.e.  $p(C) > 0$ ), knowing that  $p(C|M) = 0$  for certain plaintexts  $M$  allows the attacker to eliminate these plaintexts  $M$  from consideration (e.g.  $p(1|b) = 0$  in example.)

## Necessary Condition for Perfect Secrecy (cont'd)

Consider a cryptosystem where keys are bit strings (sequences of 0's and 1's) of some length  $k$  and messages are bit strings of some length  $m$ .

Then  $|\mathcal{K}| = 2^k$  and  $|\mathcal{M}| = 2^m$ .

The theorem shows that in order for such a system to provide perfect secrecy, we must have  $k \geq m$ , i.e. keys must be at least as long as messages!

## Shannon's Theorem

## Theorem 5 (Shannon's Theorem, 1949/50)

*A cryptosystem with  $|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$  has perfect secrecy if and only if*

- *$p(K) = 1/|\mathcal{K}|$  for all  $K \in \mathcal{K}$  (i.e. every key is chosen with equal likelihood) and*
- *for every  $M \in \mathcal{M}$  and every  $C \in \mathcal{C}$ , there exists a unique key  $K \in \mathcal{K}$  such that  $E_K(M) = C$ .*

## Proof.

See Theorem 3.4, p. 68, in Stinson-Paterson. □

## Is Perfect Secrecy the Holy Grail?

Perfect secrecy isn't all it's made out to be. For example, by Shannon's Theorem, the shift cipher — which we have seen is completely insecure — provides perfect secrecy if every key is chosen equally likely (see Theorem 3.3, pp. 66-67, of Stinson-Paterson).

We will next discuss the one-time pad, which also provides perfect secrecy but is quite impractical.

## One-Time Pad

Generally attributed to Vernam (1917, WW I) who patented it, but recent research suggests the technique may have been used as early as 1882

- in any case, it was long before Shannon

It is the only substitution cipher that does not fall to statistical analysis.

## Bitwise Exclusive-Or

Fix a string length  $n$ . Then set  $\{0, 1\}^n$  is the set of *bit strings* of length  $n$ .

### Definition 9 (bitwise exclusive or, XOR)

For  $a, b \in \{0, 1\}$ , we define

$$a \oplus b = a + b \pmod{2} = \begin{cases} 0 & a = b, \\ 1 & a \neq b. \end{cases}$$

For  $A = (a_1, a_2, \dots, a_n), B = (b_1, b_2, \dots, b_n) \in \{0, 1\}^n$ , we define then

$$A \oplus B = (a_1 \oplus b_1, a_2 \oplus b_2, \dots, a_n \oplus b_n).$$

(component-wise XOR).

## The One-Time Pad

### Definition 10 (Vernam one-time pad)

$\mathcal{M} = \mathcal{C} = \mathcal{K} = \{0, 1\}^n$  ( $n \in \mathbb{N}$ ).

Encryption of  $M \in \{0, 1\}^n$  under key  $K \in \{0, 1\}^n$  is bitwise XOR, i.e.

$$C = M \oplus K.$$

Decryption of  $C$  under  $K$  is done the same way, i.e.  $M = C \oplus K$ .

Decryption is the inverse of encryption, since  $K \oplus K = (0, 0, \dots, 0)$  and  $M \oplus (0, 0, \dots, 0) = M$ .

## Security of the One-Time Pad

### Theorem 6

*The one-time pad provides perfect secrecy if each key is chosen with equal likelihood. Under this assumption, each ciphertext occurs with equal likelihood (regardless of the probability distribution on the plaintext space).*

### Proof sketch

The first assertion follows immediately from Shannon's Theorem (Theorem 5). The second assertion is proved by computing  $p(C)$  for all  $C \in \mathcal{C}$  using the formula.  $\square$

This means that in the one-time pad, any given ciphertext can be decrypted to *any* plaintext with equal likelihood (def'n of perfect secrecy). There is no “distinguished” (e.g. meaningful) decryption. So even exhaustive search doesn't help.