CS458: Introduction to Information Security

Notes 2: Historical Crypto

Yousef M. Elmehdwi

Department of Computer Science

Illinois Institute of Technology

yelmehdwi@iit.edu

August 29th 2022

Slides: Adopted from Ewa Syta, Yale University, Computer Security: Principles and Practice, 4th Edition & Ian Goldberg, Florian Kerschbaum University of Waterloo

Outline

- Cryptography
- Cryptanalysis
- Classical Cryptography
  - Substitution Ciphers
  - Transposition Ciphers

Crypto

The basic terminology of crypto includes the following:

- Cryptography: ("secret writing"): Making secret messages
  - Turning plaintext (an ordinary readable message) into ciphertext (secret messages that are "hard" to read)
  - i.e., the science of secret writing with the goal of hiding the meaning of a message
- Cryptanalysis: Breaking secret messages
  - Recovering the plaintext from the ciphertext without knowledge of the key.
  - Also called code breaking
- Cryptology: The art and science of making and breaking "secret codes"
- Crypto: a synonym for any or all of the above (and more), where the precise meaning should be clear from context
- The point of cryptography is to send secure messages over an insecure medium (like the Internet)

Characterizing Cryptographic Systems

- Cryptographic systems are generally classified along three independent dimensions:
  - Type of operations used for encryption
    - Substitution: each element in the plaintext is mapped into another element
    - Transposition: elements in plaintext are rearranged
    - Product systems/ciphers: multiple stages/combinations of substitutions and transpositions
  - Number of keys used
    - Symmetric: sender and receiver use same key
    - Asymmetric: sender and receiver each use a different key
  - Processing of plaintext
    - Block cipher: process the input and block of elements at a time, producing output block for each input block.
    - Stream cipher: process the input elements continuously, producing output element one at a time, as it goes along.
- Key
  - Some critical information used by the cipher, known only to the sender and receiver

Terminology

The basic terminology of crypto includes the following:

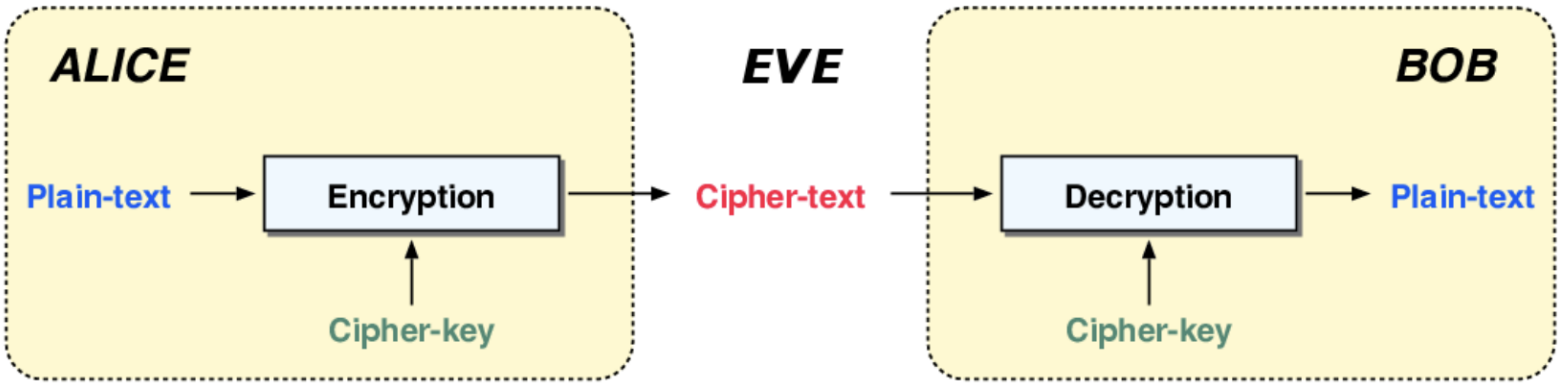
- Cipher: a particular algorithm (cryptographic system)
- Plaintext: original message
- Ciphertext: encrypted or coded message
- Encryption: convert from plaintext to ciphertext (enciphering)
- Decryption: restore the plaintext from ciphertext (deciphering)
- Key: critical information used in cipher known only to sender/receiver
- Symmetric key cryptosystem uses the same key to encrypt as to decrypt
- Public/Asymmetric key cryptosystem uses a public key to encrypt and a private key to decrypt

Secret Message Transmission Problem

- Alice wants to send Bob a private message **m** over the internet.
- Eve is an eavesdropper who listens in and wants to learn **m**.
- Alice and Bob want **m** to remain private and unknown to Eve.

Solution using encryption[1]

- A symmetric cryptosystem (sometimes called a private-key or one-key system) is a pair of efficiently-computable functions `E` and `D` such that
  - `E(k,m)` encrypts plaintext message `m` using key `k` to produce a ciphertext `c`.
  - `D(k,c)` decrypts ciphertext `c` using `k` to produce a message `m`.
- **Requirements**:
  - Correctness `D(k,E(k,m)) = m` for all keys `k` and all messages `m`.
  - Security Given `c = E(k,m)`, it is hard to find `m` without knowing `k`



[1] image credit: Derived from https://iis-people.ee.ethz.ch/~kgf/acacia/fig/alice_bob.png

Symmetric Cryptosystem Components

- Plaintext (m): original message
- Ciphertext (c): encrypted message
- Key (k): private information
- Encryption algorithm: `c = E(k,m) = E`$_k$`(m)`
- Decryption algorithm: `m = D(k,c) = D`$_k$`(c)`

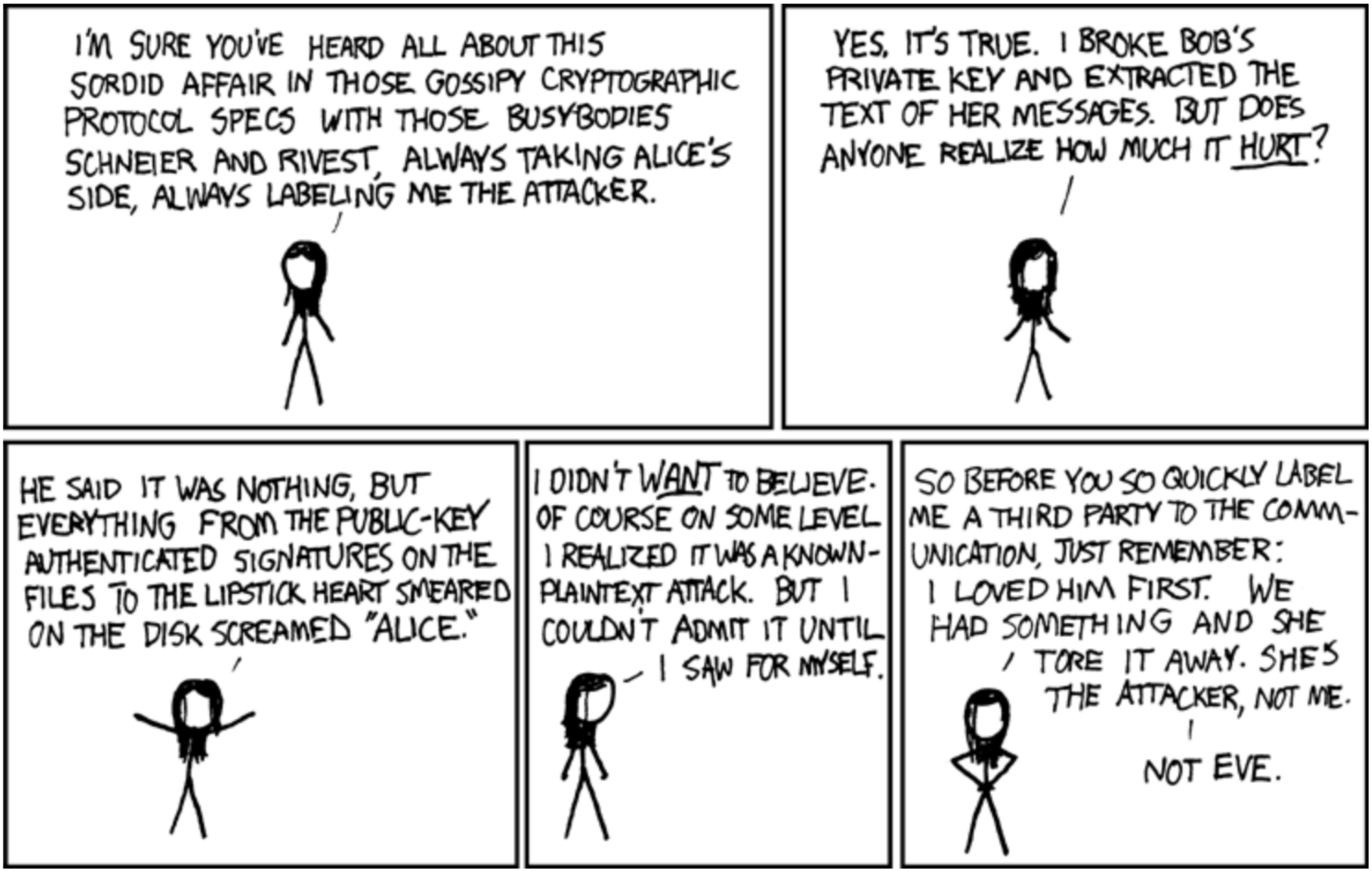
The protocol

- Protocol: Composition and ordering of the exchanged messages
- **Protocol**
  1. Alice and Bob share a common secret key k.
  2. Alice computes $c = E_k(m)$ and sends c to Bob.
  3. Bob receives c'. Computes $m' = D_k(c')$, and assumes m' to be Alice's message.
- **Assumptions**
  - Eve learns nothing except for c during the protocol.
  - The channel is perfect, so $c' = c$.
  - Eve is a passive eavesdropper who can read c but not modify it.

Requirements

- What do we require of `E`, `D`, and the computing environment?
  - Given `c`, it is hard to find `m` without also knowing `k`.
  - `k` is not initially known to Eve.
  - Eve can guess `k` with at most negligible success probability.
    - `k` must be chosen randomly from a large key space.
  - Alice and Bob successfully keep `k` secret.
    - Their computers have not been compromised; Eve can't find `k` on their computers even if she is a legitimate user, etc.
  - Eve can't obtain `k` in other ways, e.g., by social engineering, using binoculars to watch Alice or Bob's keyboard, etc.

Eve's side of the story [2]



[2] image credit: https://xkcd.com/177/

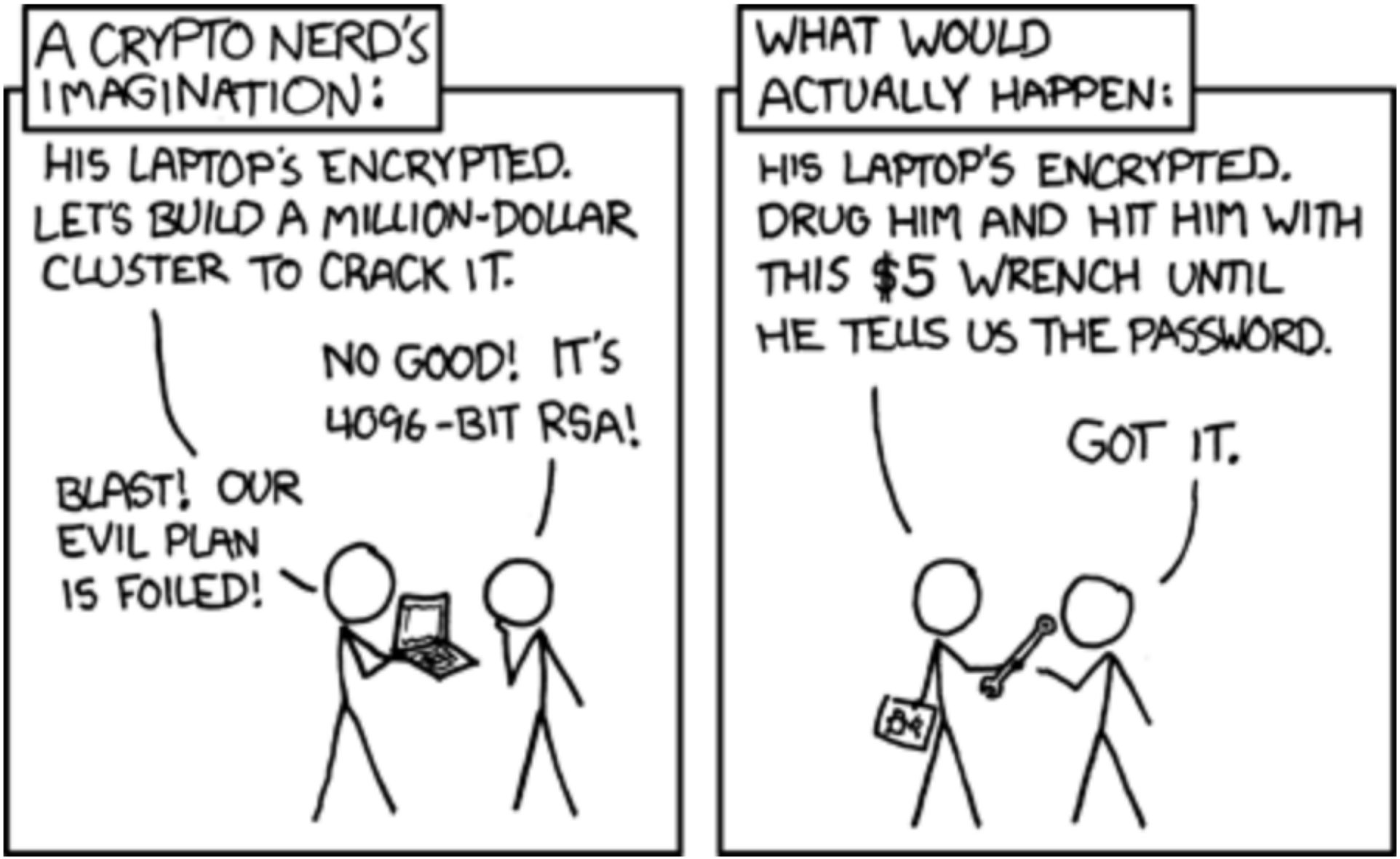
Crypto

- Basic assumptions
  - The system is completely known to the attacker
  - Only the key is secret
  - That is, crypto algorithms are not secret
- This is known as Kerckhoffs' Principle
  - i.e., one should always assume that the adversary knows the encryption/decryption algorithms and the resistance of the cipher to attacks must be based on only the secrecy of the key.
  - This has a number of implications
    - The system is at most as secure as the number of keys.
    - Eve can just try them all, until she finds the right one.
    - A strong cryptosystem is one where that's the best Eva can do.
      - with weaker systems, there are shortcuts to finding the key.
- Why do we make such an assumption?
  - Experience has shown that secret algorithms tend to be weak when exposed
  - Secret algorithms never remain secret
  - Better to find weaknesses beforehand

Eve's goals

- Eve wants learn something. Eve is not bound by any rules. She can do as she wishes with the information she has available.
- We don't want her to be able to:
  - Recover the key.
  - Find the plaintext to a ciphertext.
  - Determine any character to the plaintext.
  - Derive any meaningful information about the plaintext.

Security[3]



[3] image credit: https://xkcd.com/538/

Eve's information

- We assumed that Eve has no information about the cryptosystem except for the encryption/decryption methods and the ciphertext `c`.
- In practice, Eve might know much more.
  - She probably knows (or has a good idea) of the message distribution.
  - She might have obtained several other ciphertexts.
  - She might have learned the decryptions of earlier ciphertexts.
  - She might have even chosen the earlier messages or ciphertexts herself
- This leads us to consider several attack scenarios.

Attack scenarios

- Cryptanalysis: Process of attempting to discover the plaintext or key.
- Ciphertext-only attack
  - Eve knows only the ciphertext to be decoded c and tries to recover m.
- Known plaintext attack
  - Eve knows the ciphertext to be decoded c and a sequence of plaintext-ciphertext pairs (m_1, c_1),...,(m_r, c_r) where $c \notin \{c_1, \ldots, c_r\}$.
  - She tries to recover m.

Known plaintext attacks

- A known plaintext attack can occur when
  1. Alice uses the same key to encrypt several messages;
  2. Eve later learns or successfully guesses the corresponding plaintexts.
- Some ways that Eve learns plaintexts.
  - The plaintext might be publicly revealed at a later time, e.g., sealed bid auctions.
  - The plaintext might be guessable, e.g., an email header.
  - Eve might later discover the decrypted message on Bob's computer.

Chosen text attack scenarios

- Still stronger attack scenarios allow Eve to choose one element of a plaintext-ciphertext pair and obtain the other
- Chosen plaintext attack
  - Like a known plaintext attack, except that Eve chooses messages $m_1, \ldots, m_r$ before getting c and Alice (or Bob) encrypts them for her.
- Chosen ciphertext attack
  - Like a known plaintext attack, except that Eve chooses ciphertexts $c_1, \ldots, c_r$ before getting c and Alice (or Bob) decrypts them for her.
- Mixed chosen plaintext/chosen ciphertext attack
  - Eve chooses some plaintexts and some ciphertexts and gets the corresponding decryptions or encryptions.

Why would Alice cooperate in a chosen plaintext attack?

- Eve might be authorized to generate messages that are then encrypted and sent to Bob, but she isn't authorized to read other people's messages.[4]
- Alice might be an internet server, not a person, that encrypts messages received in the course of carrying out a more complicated cryptographic protocol.
- Eve might gain access to Alice's computer, perhaps only for a short time, when Alice steps away from her desk.

[4] Nothing we have said implies that Eve is unknown to Alice and Bob or that she is not also a legitimate participant in the protocol.

Adaptive chosen text attack scenarios

- Adaptive versions of chosen text protocols are when Eve chooses her texts one at a time after learning the response to her previous text
- Adaptive chosen plaintext attack
  - Eve chooses the messages $m_1, m_2, \ldots$ one at a time rather than all at once.
  - Thus,
    - m_2 depends on (m_1, c_1)
    - m_3 depends on both (m_1, c_1) and (m_2, c_2), etc.
- Adaptive chosen ciphertext and adaptive mixed attacks
  - are defined similarly

Computationally Secure Encryption Schemes

- Only relatively weak algorithms fail to withstand a ciphertext-only attack.
- Generally, an encryption algorithm is designed to withstand a known-plaintext attack
- Most cryptosystems have "computationally" security
  - This means that it's certain they can be broken, given enough work by Eve.
- How much is "enough"?
  - At worst, Eve tries every key
    - How long that takes depends on how long the keys are.
    - But, it only takes this long if there are no shortcuts.

Computationally Secure Encryption Schemes

- Encryption is computationally secure if either:
  - Cost of breaking cipher exceeds value of information
  - Time required to break cipher exceeds the useful lifetime of the information
- Hard to estimate value/lifetime of some information
- Harder to estimate how much effort needed to break a cipher
- Can estimate time/cost of a brute-force attack (Exhaustive Key Search)

Attacks

- Goal of the Attacker
  - Discover the plaintext (good)
  - Discover the key (better)
- Assumed Attacker Knowledge
  - Ciphertext
  - Algorithm
  - Other pairs of (plaintext, ciphertext) using same key
- Attack Methods
  - Brute-force attack
    - Try every possible key on ciphertext until readable text is obtained from the ciphertext
    - On average, number of guesses is half the key space
  - Cryptanalysis
    - Use knowledge of algorithm and/or plaintext patterns to "ntelligently" decipher the ciphertext
    - Exploit characteristics of algorithm to deduce plaintext or key
    - Attacks differ based on amount of information known to attacker
- Assumption: attacker can recognize correct plaintext
- *Cryptanalyst or attacker tries to break the system*

Exhaustive Key Search

- Exhaustive key search
  - Eve can simply try all possible keys and test each to see if it is correct.
  - Remember, she has some ciphertexts so she knows when she found the right key.
- To prevent an exhaustive key search, a cryptosystem must have a large keyspace.
  - The set of all possible keys that can be used to generate a key.
  - Must be too many keys for Eve to try them all in any reasonable amount of time.

Beyond Exhaustive Search

- A large keyspace is necessary for security.
- But a large keyspace is not sufficient.
  - Shortcut attacks might exist.
  - In cryptography, we can (almost) never prove that no shortcut attack exists

- For now, we will focus on classical crypto.
- All classical ciphers are symmetric.

Substitution Ciphers

- Probably the most common form of cipher.
- They work by replacing each letter of the plaintext with another letter (or possibly even a random symbol).
- The most famous one comes after the Caesar cipher.

Caesar Cipher/shift

- **Idea**: substitute one letter for another one.
- Julius Caesar shifts each letter by three positions in the alphabet:
- Plaintext: fourscoreandsevenyearsago
- Key: how we substitute

Plain	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
Cipher	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

- Ciphertext: IRXUVFRUHDQGVHYHQBHDUVDJR
- *Replace each letter with one 3 letters later in the alphabet.*
- The replacement remains the same throughout the message.
- The cipher is classified as a type of `Mono-alphabetic substitution` (fixed replacement structure)
  - Same letter is replaced with only one other (always the same for given cipher message).

Caesar's Cipher Decryption

- Caesar code decryption
  - Deciphering is done in reverse, with a left shift of 3.
  - Replaces a letter another with an inverse alphabet shift: a previous letter in the alphabet.
- Here is a Caesar's cipher using a right shift of three places

Plain	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
Cipher	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

- Given ciphertext:

  JHW D YDFFLQH, ZHDU D PDVN, VRFLDL SKBVLFDO GLVWDQFH

- Plaintext:

  get a vaccine, wear a mask, social distance

Shift Cipher

- We can shift by any number of positions:
  - shift by k for some $k \in \{0, 1, 2, \ldots, 25\}$
- Then, key is k
- Example: key $k = 7$.

Plain	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
Cipher	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G

Shift Cipher Encryption/Decryption

- Another way to encrypt/decrypt, more mathematical
  - Let, M: plaintext; K: key; E: encryption function; D: decryption function
  - M = {sequences of letters}
  - The encryption can also be represented using modular arithmetic by first transforming the letters into numbers, according to the scheme,
    - A $\rightarrow$0, B $\rightarrow$1, ..., Z $\rightarrow$25
  - Encryption of a letter m by a shift k can be described mathematically as
    - $K=\{k \mid k \text{ is an integer and } 0 \leq k \leq 25\}$
    - $E=\{E_k \mid k \in K,\ E_k(m)=(m+k) \bmod 26\}$
  - Decryption of a letter c by a shift k can be described mathematically as
    - $D=\{D_k \mid k \in K,\ D_k(c)=(26+c-k) \bmod 26\}$

Breaking the Shift Cipher

- Cryptanalysis attempts to discover the key or the plaintext of an encrypted message
- Imagine you have the ciphertext. How to find the key?
- A simple substitution (shift by k) is used.
  - But the key is unknown
- Given ciphertext: CSYEVIXIVQMREXIH
- Exhaustive key search
  - If the key space is small enough, try them all approach.
  - Only 26 possible keys.
  - Solution:
    - key is $k = 4$
    - Plaintext: youareterminated

Permutation Cipher

- In general, simple substitution key can be any permutation of letters
- This is a simple substitution cipher in which the mapping from plaintext alphabet to ciphertext alphabet is an arbitrary permutation of the letters of the alphabet.
  - The secret-key is a permutation π from `{0..25}` to `{0..25}` drawn uniformly at random
  - Not necessarily a shift of the alphabet.
- How many keys are possible?
- For example:

Plain	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
Cipher	J	I	C	A	X	S	E	Y	V	D	K	W	B	Q	T	Z	R	H	F	M	P	N	U	L	G	O

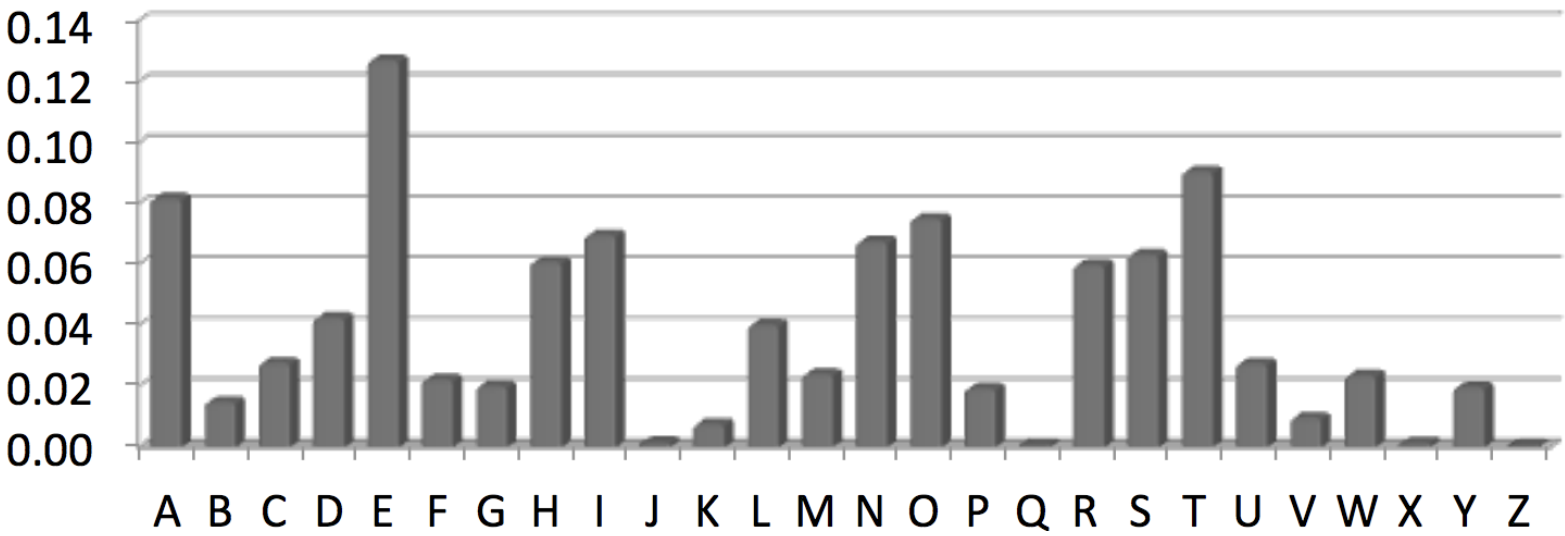
- Then $26 \times 25 \times 24 \times 23 \ldots \times 3 \times 2 \times 1 = 26! = 403291461126605635584000000$ possible keys!
  - years to try every key

Cryptanalysis: Be Clever

- Cannot try all `26!` simple substitution keys.
- Can we be more clever?
- But knowledge of language statistics makes it easy to break
  - If attacker knows the message is in plain English can use known patterns in English language:
    - Frequency of letters
    - Frequency of pairs of letters (digrams) and triples of letters (trigrams)
    - Known or expected words in plaintext
  - *Break it by guessing words from their letter patterns, or by using the relative frequency of individual letters.* `E` *is most frequent,* `T` *is second, etc.*

Cryptanalysis: Frequency Analysis

- Frequency analysis is the study of the frequency of letters or groups of letters in a ciphertext.
- Frequency analysis is a technique based on the fact that, in any given stretch of written language, certain letters and combinations of letters occur with varying frequencies.
  - Some letters more popular than others.
  - Some pairs of letters more popular than others
- The following chart shows the frequency of each letter of the alphabet for the English language



Cryptanalysis II: frequency analysis

- Ciphertext:

```
D  RNXHT  VHRVCK  VKKXOW  FYVF  V  OVFY

GENBWKKNE'K  PWEC  BVPNEDFW  TWKKWEF  DK  GD.
```

- Simple substitution
- No letter is encrypted as itself.
  - For example, in this message we know that `PWEC` cannot be the ciphertext for `when`.
- Analyze this message

Cryptanalysis of substitution ciphers

- Frequency analysis works well with substitution ciphers.
- We replace one letter with another one but it doesn't affect the frequency distributions.
  - Calculate the frequency table.
  - Try to guess the most popular letters.
  - Try to find pairs and triples of letters.
  - Fill in the blanks.

Transposition

- Let's try another approach to hide information.
- What else can we do with the plaintext message?
- Instead of replacing letters, focus on their positions.
- Transposition ciphers
  - The name given to any encryption that involves rearranging the plaintext letters in a new order
  - i.e., rearrange characters of plaintext to produce ciphertext.

Simple Transposition

- The key to the cipher is the number of rows and columns of a matrix.
- Encipher a message by writing the plaintext into the matrix by rows and reading the ciphertext out of the matrix by columns.
- Plaintext: attackxatxdawn

a	t	t	a
c	k	x	a
t	x	d	a
w	n	x	x

- Ciphertext: ACTWTKXNTXDXAAAX

Double Transposition

- Can we do better?
- We could put the plaintext into an array and permute the rows and columns
- Plaintext: attackatdawn

$$\begin{bmatrix} a & t & t & a \\ c & k & a & t \\ d & a & w & n \end{bmatrix} \longrightarrow \begin{bmatrix} d & a & w & n \\ c & k & a & t \\ a & t & t & a \end{bmatrix} \longrightarrow \begin{bmatrix} n & a & d & w \\ t & k & c & a \\ a & t & a & t \end{bmatrix}$$

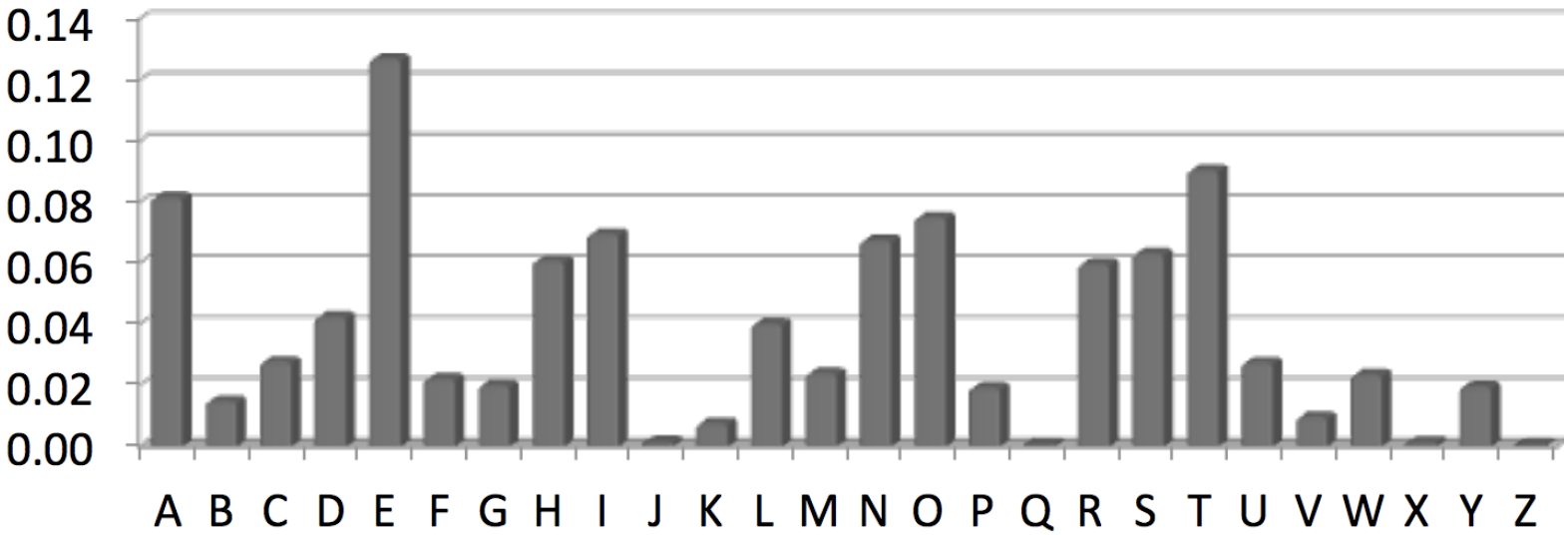
- Key is the matrix size and the row and column permutations: `(3,2,1)` and `(4,2,1,3)`.
- Ciphertext: NADWTKCAATAT

Cryptanalysis

- You are given the ciphertext xtawxnattxadakc. How do you find the plaintext?
- Assume you know a transposition cipher was used.
  - You need to reconstruct the matrix and figure out the scrambling method.
  - Single transposition: guess the number of columns.
  - Double transposition: also need the column and row ordering.

Cryptanalysis

- We learned about frequency analysis! Why can't we use it here?
- Well, we can. But will it do us any good?
- This is what you will get.



- Q: What is going on here?

Big Crypto Ideas

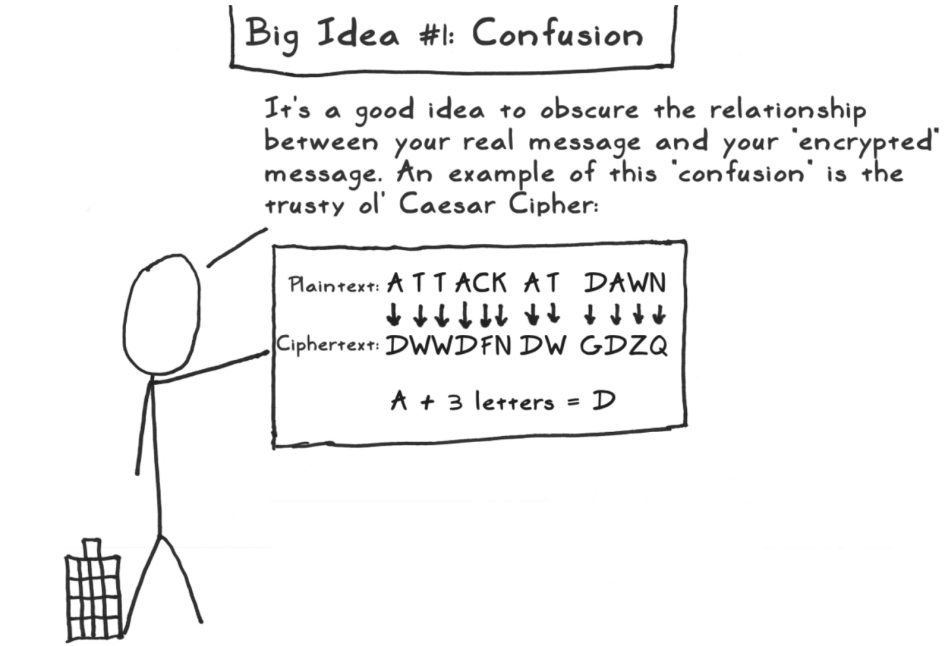
- So, what have we learned so far?
- **3 Big Ideas**:
  - Big Idea #1: Confusion
  - Big Idea #2: Diffusion
  - Big Idea #3: Key secrecy

Confusion & Diffusion

- Confusion and Diffusion are two properties of the operation of a secure cipher which were identified by Claude Shannon in his paper *Communication Theory of Secrecy Systems*[5].
- DES, AES and many block ciphers are designed using Shannon's idea of confusion and diffusion.

[5] http://pages.cs.wisc.edu/~rist/642-spring-2014/shannon-secrecy.pdf

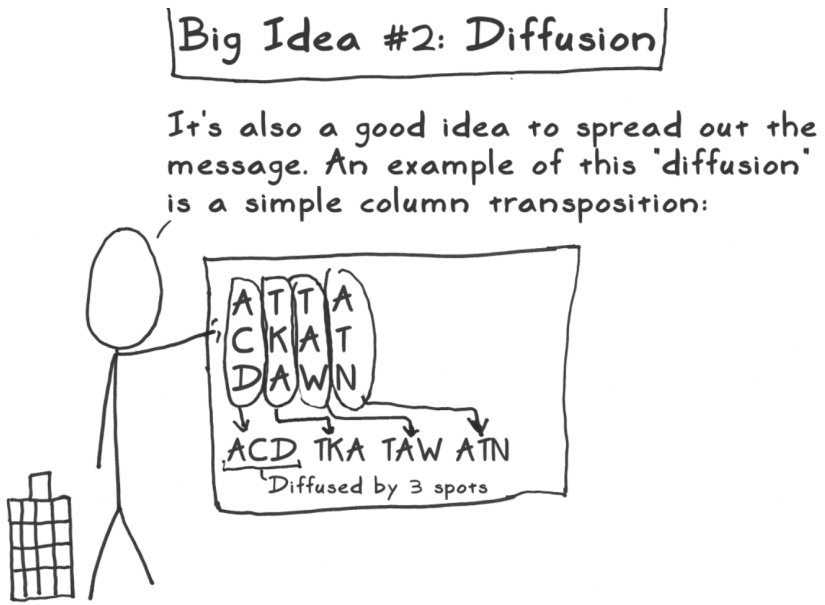
Confusion



www.moserware.com/2009/09/stick-figure-guide-to-advanced.html

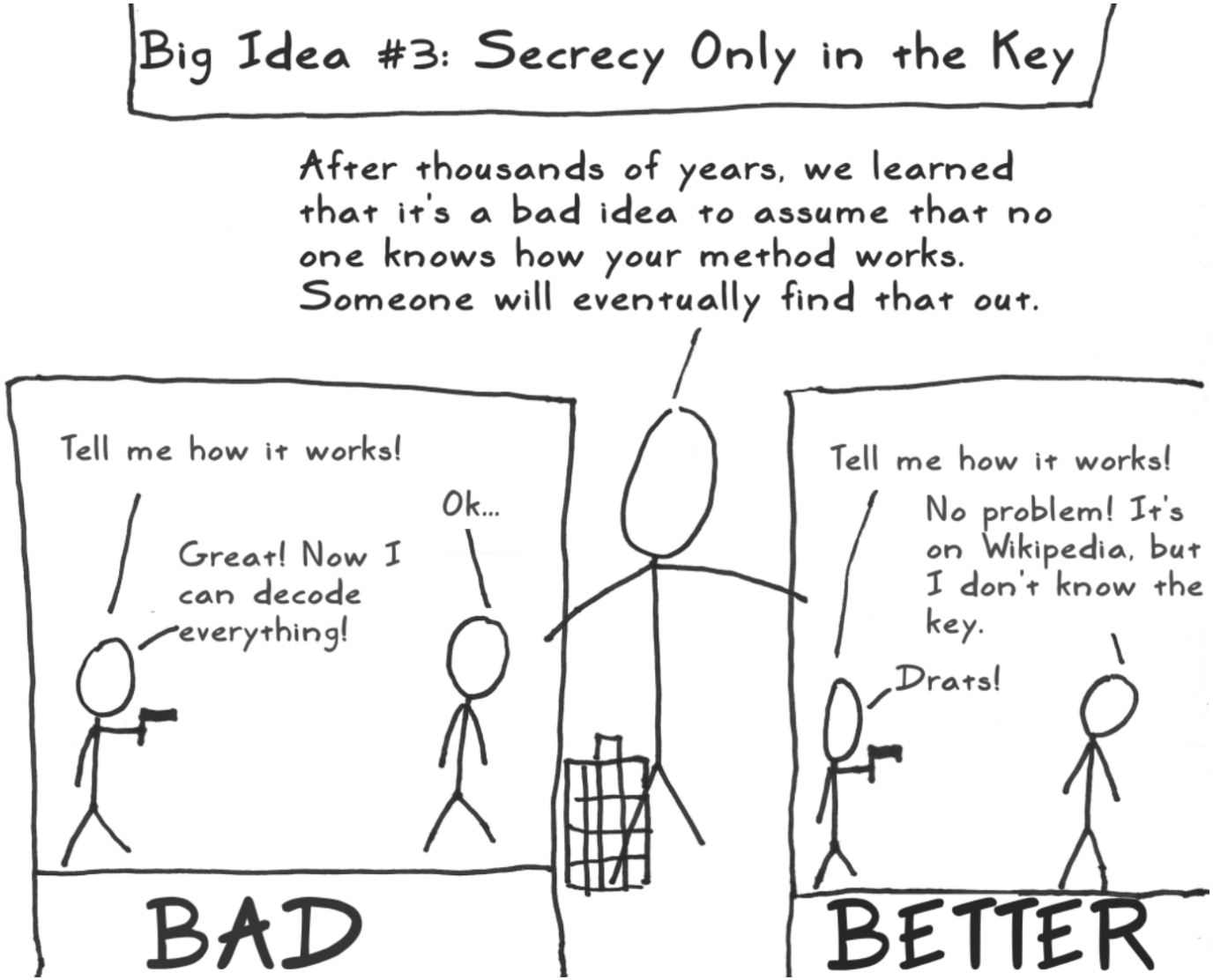
- Confusion: It is a technique of ensuring that a ciphertext gives no clue about plaintext.
- The property of confusion
  - hide the relationship between the ciphertext and the key.
  - make it difficult to find the key from the ciphertext
- *No leaks of information. You should not be able to find patterns*

Diffusion



www.moserware.com/2009/09/stick-figure-guide-to-advanced.html

- The idea of diffusion is to hide the relationship between the ciphertext and the plaintext
- If we change a character in the plaintext, how that should affect the ciphertext
  - *a little change in the plaintext, should result a big change in the ciphertext*



www.moserware.com/2009/09/stick-figure-guide-to-advanced.html

This is known as Kerckhoffs Principle.

Combining Ciphers

- Confusion (substitution) and diffusion (transposition) on their own are not enough.
- What if we combine multiple substitution **or** multiple transposition ciphers?
  - Two (or more) substitutions are really only one more complex substitution.
  - Two (or more) transpositions are really only one more complex transposition.
- But: it makes sense to combine substitution and transposition!
- You get the best of both worlds!

Mono-alphabetic cipher

- A shift cipher uses a letter substitution defined by a rotation of the alphabet.
- Any cipher that uses a substitution to replace a plaintext letter by a ciphertext letter is called a substitution cipher.
  - A shift cipher is a special case of a substitution cipher.
- Any cipher that encrypts a message by applying the same substitution to each letter of the message is called a mono-alphabetic cipher.
- A mono-alphabetic substitution cipher, also known as a simple substitution cipher, relies on a fixed replacement structure. That is, the substitution is fixed for each letter of the alphabet. Thus, if "a" is encrypted to "S", then every time we see the letter "a" in the plaintext, we replace it with the letter "S" in the ciphertext.

Polyalphabetic ciphers

- Another way to strengthen substitution ciphers is to use different substitutions for different letter positions.
  - Choose r different alphabet permutations $\pi_1, \ldots, \pi_r$ for some number r.
  - Use π_1 for the first letter of m, π_2 for the second letter, etc.
  - Repeat this sequence after every r letters.
- While this is much harder to break than monoalphabetic ciphers, letter frequency analysis can still be used.
- Every r^{th} letter is encrypted using the same permutation, so the submessage consisting of just those letters still exhibits normal English language letter frequencies

Vigènere Cipher

- The Vigenère cipher is a polyalphabetic cipher in which the number of different substitutions r is also part of the key.
- Thus, the adversary must determine r as well as discover the different substitutions.
- All polyalphabetic ciphers can be broken using letter frequency analysis, but they are secure enough against manual attacks to have been used at various times in the past.
- The German Enigma encryption machine used in the second world war is also based on a polyalphabetic cipher.

Vigènere Cipher

- Like Caesar cipher, but use a phrase as key
- Example
  - `Message`: THE BOY HAS THE BALL
  - `Key`: VIG
  - `Encipher`: using shift cipher for each letter:

```
key     VIGVIGVIGVIGVIGV
plain   THEBOYHASTHEBALL
cipher  OPKWWECIYOPKWIRG
```

- The Vigenère cipher uses a 26×26 table with A to Z as the row heading and column heading.
- This table is usually referred to as the Vigenère Tableau, Vigenère Table or Vigenère Square.

The Vigènere Tableau

	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*
A	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
B	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	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	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

How this Cipher Works

1. Pick a keyword (for our example, the keyword will be `VIG`).
2. Write the keyword across the top of the text you want to encipher, repeating it as many times as necessary.
3. For each letter, look at the letter of the keyword above it (if it was 'V', then you would go to the row that starts with an 'V'), and find that row in the Vigenère table.
4. Then find the column of your plaintext letter (for example, 't', so the twenty-th column).
5. Finally, trace down that column until you reach the row you found before and write down the letter in the cell where they intersect (in this case, you find an 'o' there).

Vigènere Cipher

	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*
A	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
B	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	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	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

- Example
- key V, letter T: follow V column down to T row (giving “O”)
- key I, letter H: follow I column down to H row (giving “P”)

```
key     VIGVIGVIGVIGVIGV
plain   THEBOYHASTHEBALL
cipher  OPKWWECIYOPKWIRG
```

Decrypt

1. Pick a keyword (for our example, the keyword will be `VIG`).
2. Write the keyword across the top of the text you want to decrypt, repeating it as many times as necessary.
3. For each letter in the ciphertext, look at the letter of the keyword above it (if it was 'V', then you would go to the row that starts with an 'V'), and find that row in the Vigenère table.
4. Then, this row is used to find the letter heading of that column that contains the ciphertext letter is the needed plaintext letter (in this case, you find an 't' there).

Useful Terms

- period: length of key
  - In earlier example, period is `3`
- tableau: table used to encipher and decipher
  - Vigènere cipher has plaintext letters on top, key letters on the left
- polyalphabetic: the key has several different letters
  - Caesar cipher is mono-alphabetic

Security of Vigènere Cipher

- Vigènere masks the frequency with which a character appears in a language:
  - one letter in the ciphertext corresponds to multiple letters in the plaintext.
- Makes the use of frequency analysis more difficult.
- Any message encrypted by a Vigènere cipher is a collection of as many shift ciphers as there are letters in the key.

Attacking the Cipher

- Approach
  - Establish period (find the length of the key); call it r
  - Break message into r parts, each part being enciphered using the same key letter
  - Solve each part

Cryptanalysis: Terminology

- Cryptosystem is secure if best known attack is to try all keys.
  - Exhaustive key search, that is.
- Cryptosystem is insecure if any shortcut attack is known.
- Q: Are there any completely secure ciphers?

Perfect Secret-key Encryption

- Is it possible to make a completely unbreakable cryptosystem?
- Yes: The Vernam cipher (one-time pad)

Vernam cipher (One-time pad)

- One-time pad is a cipher that cannot be broken if it is used correctly.
- Rules:
  - The key is as long as the message.
  - The key is random.
  - The key is never reused.

Exclusive-or on bits

- Vernam cipher is based on exclusive-or (XOR), which we write as $\oplus$
  - $x \oplus y$ is true when exactly one of x and y is true.
  - $x \oplus y$ is false when x and y are both true or both false.
- Exclusive-or is just sum modulo two if 1 represents true and 0 represents false.

$$x \oplus y = (x + y) \bmod 2$$

- XOR is associative and commutative. 0 is the identity element.

$$k \oplus 0 = 0 \oplus k = k$$

- XOR is its own inverse.

$$k \oplus k = 0$$

One-Time Pad: Informal description

- The one-time pad encrypts a message m by XORing it with the key k, which must be as long as m.
- Assume both m and k are represented by strings of bits. Then ciphertext bit $c_i = m_i \oplus k_i$.
- Note that $c_i = m_i$ if $k_i = 0$, and $c_i = \neg m_i$ if $k_i = 1$.
- Decryption is the same, i.e., $m_i = c_i \oplus k_i$

One-Time Pad: Encryption

- Let a=000, h=001, i=011, k=100, p=101, y=111
- **Encryption:** Plaintext $\oplus$ Key = Ciphertext

	h	a	p	p	y
Plaintext	001	000	101	101	111
Key	101	111	110	101	011
Ciphertext	100	111	011	000	100
	k	y	i	a	k

One-Time Pad: Decryption

- Let a=000, h=001, i=011, k=100, p=101, y=111
- **Decryption:** Ciphertext $\oplus$ Key = Plaintext

	k	y	i	a	k
Ciphertext	100	111	011	000	100
Key	101	111	110	101	011
Plaintext	001	000	101	101	111
	h	a	p	p	y

The one-time pad cryptosystem formally defined

- $\mathcal{M} = \mathcal{C} = \mathcal{K} = \{0,1\}^r$ for some length r.
- $E_k(m) = k \oplus m$, where $\oplus$ is applied to corresponding bits of k & m.
- $D_k(c) = k \oplus c$, where $\oplus$ is applied to corresponding bits of k & c.
- It works because

$$D_k(E_k(m)) = k \oplus (k \oplus m) = (k \oplus k) \oplus m = 0 \oplus m = m$$

One-time pad: Security

- Like the 1-letter Caesar cipher, for given m and c, there is exactly one key k such that $E_k(m) = c$ (namely, $k = m \oplus c$)
- For fixed c, m varies over all possible messages as k ranges over all possible keys, so c gives no information about m.
- It will follow that the one-time pad is information-theoretically secure[6]

[6] *Information-theoretic security is a cryptosystem whose security derives purely from information theory. In other words, it cannot be broken even if the adversary had unlimited computing power. The adversary simply does not have enough information to break the encryption and so the cryptosystems are considered cryptanalytically-unbreakable*

Importance of the Vernam cipher

- It is important because
  - it is sometimes used in practice;
  - it is the basis for many stream ciphers, where the truly random key is replaced by a pseudo-random bit string.

Attraction of one-time pad

- The one-time pad would seem to be the perfect cryptosystem.
  - It works for messages of any length (by choosing a key of the same length).
  - It is easy to encrypt and decrypt.
  - It is information-theoretically secure.
- In fact, it is sometimes used for highly sensitive data.

Drawbacks of one-time pad

- It has two major drawbacks:
  1. The key k must be as long as the message to be encrypted.
  2. The same key must never be used more than once. (Hence the term "one-time".)
- Together, these make the problem of key distribution and key management very difficult.

Why the key cannot be reused

- If Eve knows just one plaintext-ciphertext pair (m_1, c_1), then she can recover the key $k = m_1 \oplus c_1$.
- This allows her to decrypt all future messages sent with that key.
- Even in a ciphertext-only situation, if Eve has two ciphertexts c_1 and c_2 encrypted by the same key k, she can gain significant partial information about the corresponding messages m_1 and m_2.
  - In particular, she can compute $m_1 \oplus m_2$ without knowing either m_1 or m_2

$$c_1 \oplus c_2 = (m_1 \oplus k) \oplus (m_2 \oplus k) = m_1 \oplus m_2$$

How knowing $m_1 \oplus m_2$ might help an attacker

- Fact (important property of $\oplus$)
  - For bits b_1 and b_2, $b_1 \oplus b_2 = 0$ if and only if $b_1 = b_2$
  - Hence, blocks of `0's` in $m_1 \oplus m_2$ indicate regions where the two messages m_1 and m_2 are identical.
  - That information, together with other information, Eve might have about the likely content of the messages, may be enough for her to seriously compromise the secrecy of the data.

One-Time Pad Summary

- Ciphertext provides no info about plaintext.
- But, only when used correctly!
  - Pad must be random, used only once.
  - Pad is known only to sender and receiver.
- Note: pad (key) is same size as message.
  - So, why not distribute message instead of pad?

Key Points

- Two basic types of ciphers
  - Transposition ciphers and substitution ciphers
- Caesar cipher uses one key
- Vigènere cipher uses a sequence of keys
- One-time pad is a provably secure cryptosystem but not practical to use in modern communication
- Cryptanalysis
  - Exhaustive search
  - Statistical analysis

Reading

- Information Security: Principles and Practice, 2nd edition (available online)
  - Chapter 2 (Till 2.3.5)
- Computer Security: Principles and Practice
  - Chapter 20 (Only 20.1)

- Next: Modern Cryptography