CS458: Introduction to Information Security

Notes 3: Historical Crypto - Part I

Yousef M. Elmehdwi

Department of Computer Science

Illinois Institute of Technology

yelmehdwi@iit.edu

August 30, 2018

Slides: Adopted from Ewa Syta, Yale University

Outline

- History of Cryptography
- Classical Cryptography
  - Substitution Ciphers
  - Transposition Ciphers

- Crypto

History of Cryptography

- Contrary to popular belief, crypto has been around for a long time. It has been used for thousands of years to hide secret messages.
- However, cryptology is a young science. Systematic study of cryptology as a science just started around one hundred years ago.

History of Cryptography

- The first known evidence of the use of cryptography was found in an inscription carved around 1900 BC, in the main chamber of the tomb of the nobleman Khnumhotep II, in Egypt
- However, cryptology is a young science. Systematic study of cryptology as a science just started around one hundred years ago.

History of Cryptography[1]

- The Greeks wrapped a tape around a stick, and then wrote the message on it. When the tape was unwrap, the writing would be meaningless. The receiver of the message would need a stick of the same diameter.

- The device (scylate) was first mentioned in the 7th century BC. This is an example of a transposition.

[1] image credit: Wikipedia

History of Cryptography

- The most popular cipher is due to Julius Caesar (100 BC-44 BC). He used it to convey secret messages to his army generals posted in the war front.
- It is based on a simple substitution of letters.

Terminology

- Cryptology: The art and science of making and breaking "secret codes"
- Cryptography: making "secret codes"
- Cryptanalysis: breaking "secret codes"
- Crypto: all of the above (and more)

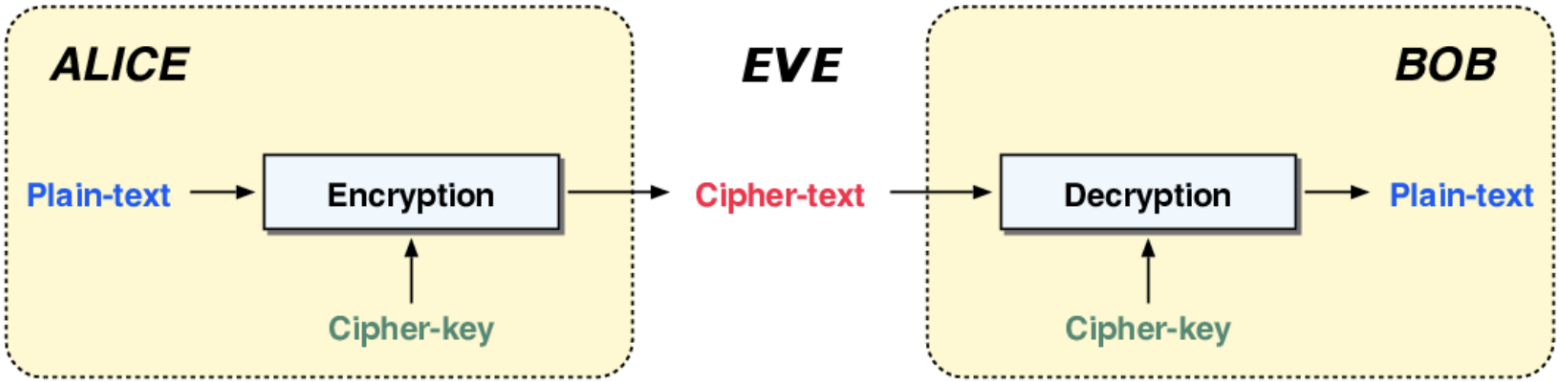
- Secret Message Transmission

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: $m = D(k,c) = D_k(c)$

Secret Message Transmission Problem [2]

- 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.



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

Solution using encryption

- 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`

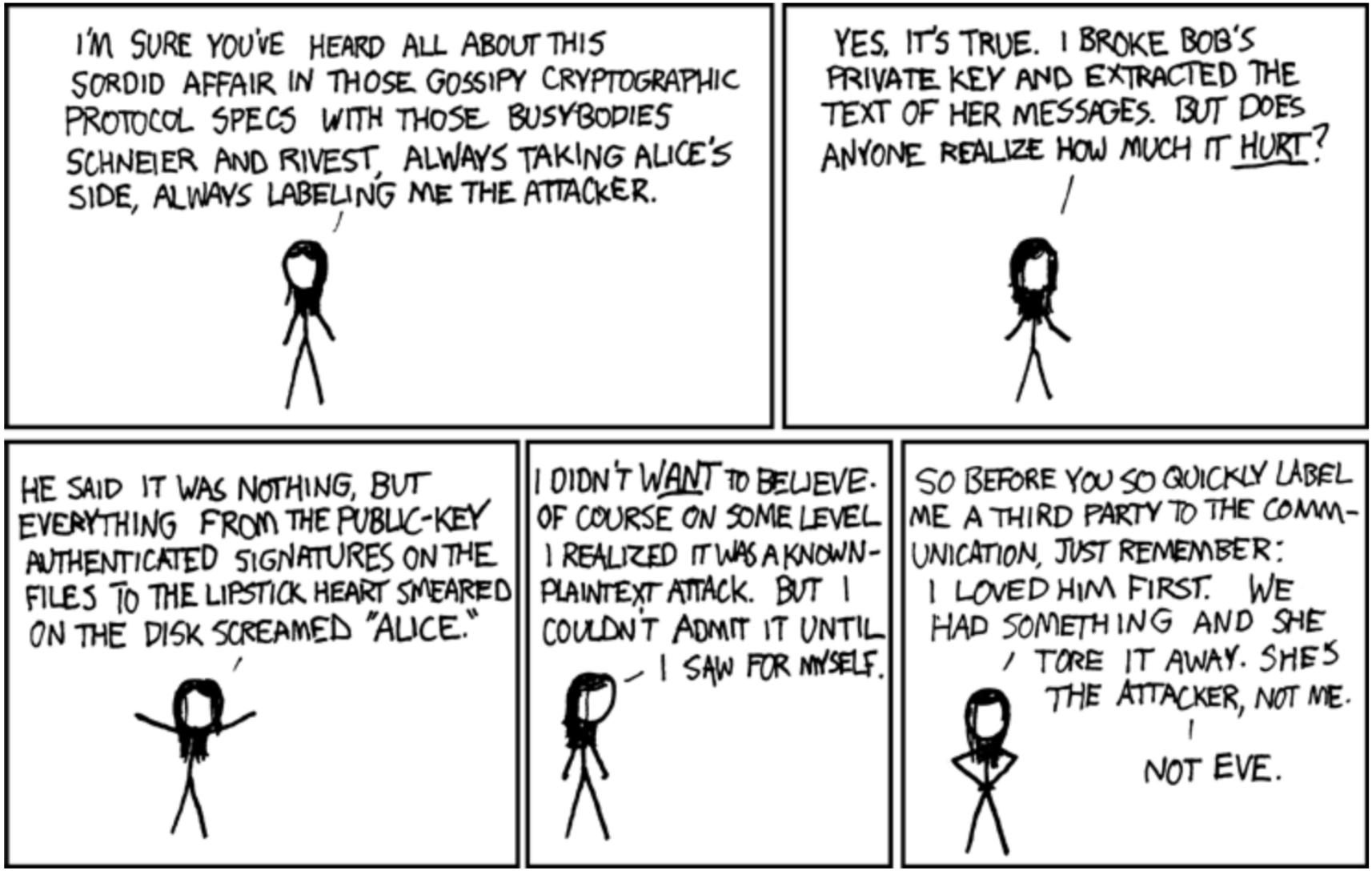
The protocol

- **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 [3]



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

Crypto

- Classical Cryptography
  - Used in the past.
  - “Pen and paper” ciphers.
  - Easily broken.
- **Modern Cryptography**
  - Symmetric crypto and public key (asymmetric) crypto
  - Strong
  - AES, DES, RSA, ElGamal
- So, why study the classical crypto methods at all?

Lessons learned from classical crypto

- While on the surface modern cryptographic techniques look nothing like the classical ones, they follow the same major principles.
- We learned about those principles from the simple ciphers.
  - Confusion
  - Diffusion
  - Key secrecy
- We will learn about these concepts.

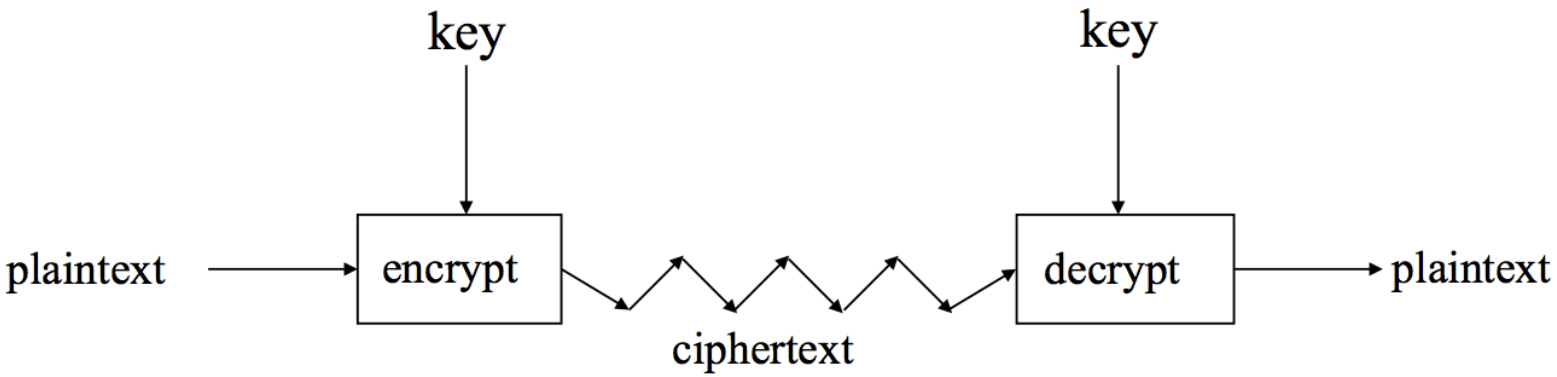
How to Speak Crypto

- A cipher or cryptosystem is used to encrypt the plaintext.
- The result of encryption is ciphertext.
- We decrypt ciphertext to recover plaintext.
- A key is used to configure a cryptosystem.
- A symmetric key cryptosystem uses the same key to encrypt as to decrypt.
- A public key or asymmetric cryptosystem uses a public key to encrypt and a private key to decrypt.

Crypto

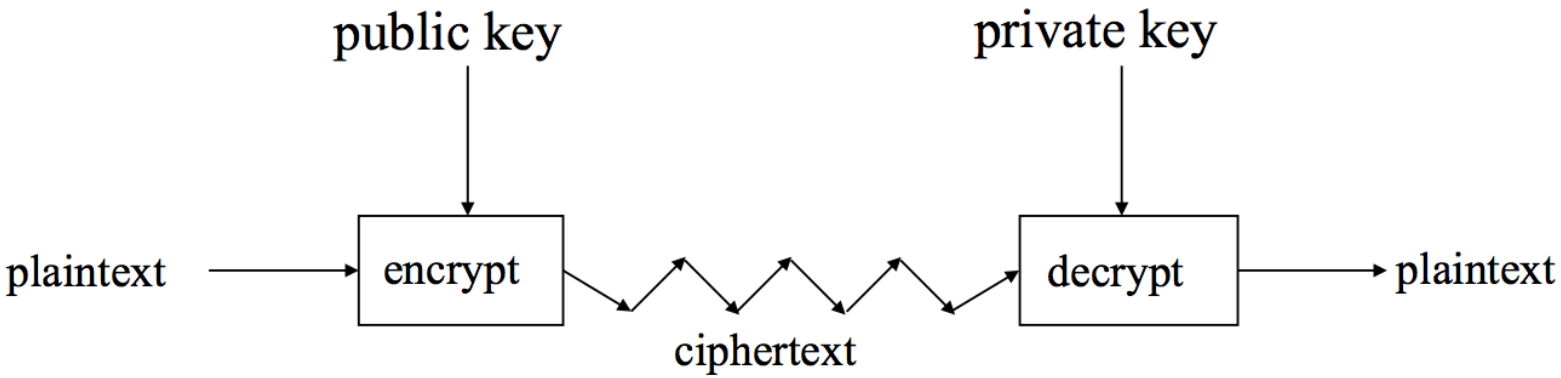
- Basic assumptions
  - The system is completely known to the attacker
  - Only the key is secret
  - That is, crypto algorithms are not secret
- This is known as Kerckhoffs' Principle
  - i.e., one should always assume that the adversary knows the encryption/decryption algorithms and the resistance of the cipher to attacks must be based on only the secrecy of the key
- 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

Crypto as Black Box: Symmetric



A generic view of symmetric key crypto

Crypto as Black Box: Asymmetric



A generic view of public key (asymmetric key) crypto

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

Simple Substitution

- **Idea**: substitute one letter for another one. But we need some order!
- Plaintext: fourscoreandsevenyearsago
- Key: how we substitute

Plaintext	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
Ciphertext	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
- Here we shift letters of the alphabet. If we shift by 3, we get the Caesar's cipher

Ceasar's Cipher Decryption

- Suppose we know a Caesar's cipher is being used:

Plaintext	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
Ciphertext	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: VSRQJHEREVTXDUHSDQWV
- Plaintext: spongebobsquarepants

Ceasar's Cipher Encryption/Decryption

- Let, M: plaintext; K: key; E: encryption function; D: decryption function
- M = {sequences of letters}
- K={i | i is an integer and $0 \leq i \leq 25$}
- E={E | $k \in K$ and for all letters m, $E_k(m)=(m+k) \bmod 26$}
- D={D| $k \in K$ and for all letters c, $D_k(c)=(26+c-k) \bmod 26$}

Not-so-Simple Substitution

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

Plaintext	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
Ciphertext	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

Cryptanalysis I: Try Them All

- 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 `n`) is used.
  - But the key is unknown
- Given ciphertext: CSYEVIXIVQMREXIH
- Exhaustive key search
  - Try them all approach.
  - Only 26 possible keys.
  - Solution: key is `n = 4`

Simple Substitution: General Case

- In general, simple substitution key can be any permutation of letters
  - Not necessarily a shift of the alphabet.
- How many keys are possible?
- For example:

Plaintext	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
Ciphertext	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

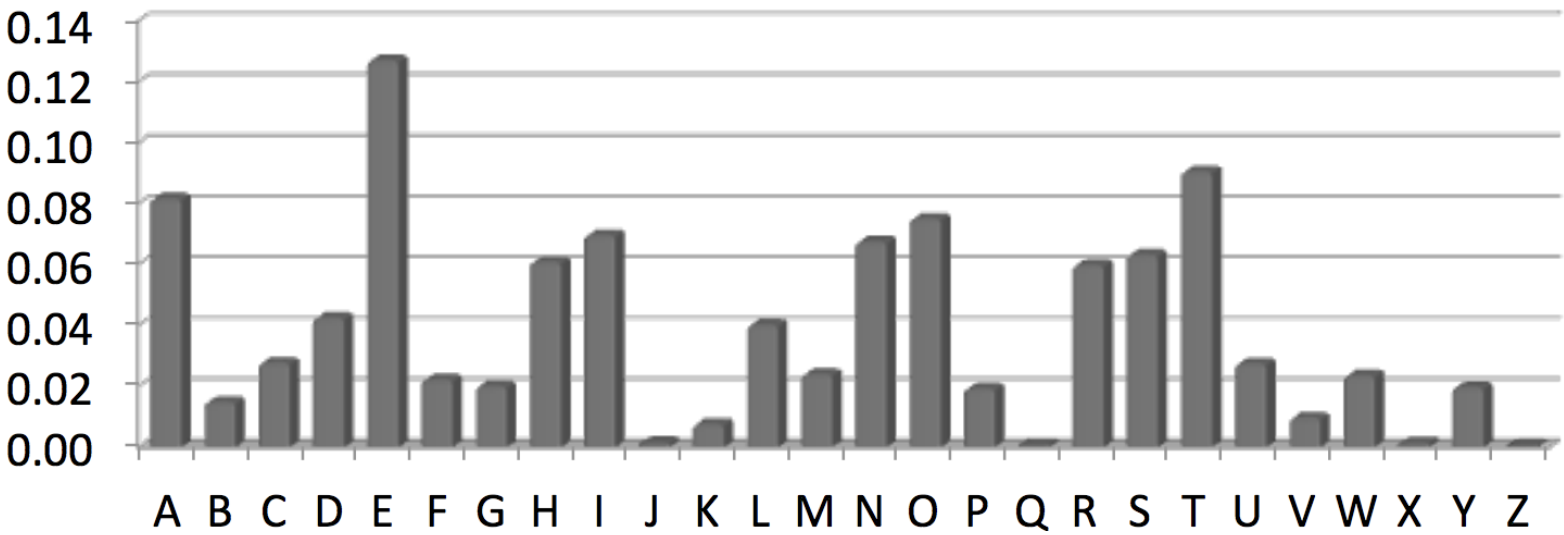
- Then $26 \times 25 \times 24 \times 23 \ldots \times 3 \times 2 \times 1 = 26!$ possible keys!

Cryptanalysis II: Be Clever

- Cannot try all simple substitution keys.
- Can we be more clever?
- What if you know the message is in English?

Cryptanalysis II: frequency analysis

- 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



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 II: frequency analysis

- Ciphertext:

PBFPVYFBQXZTYFPBFEQJHDXXQVAPTPQJKTOYQWIPBVWLXTOXBTF
XQWAXBVCXQWAXFQJVWLEQNTOZQGGQLFXQWAKVWLXQWAEBIP
BFXFQVXGTVJVWLBTPQWAEBFPBFHCVLXBQUFEVWLXGDPEQVPQG
VPPBFTIXPFHXZHVFAGFOTHFEFBQUFTDHZBQPOTHXTYFTODXQHFT
DPTOGHFQPBQWAQJJTODXQHFOQPWTBDHHIXQVAPBFZQHCFWPFHP
BFIPBQWKFABVYYDZBOTHPBQPQJTQOTOGHFQAPBFEQJHDXXQVAV
XEBQPEFZBVFOJIWFFACFCCFHQWAUVWFLQHGFXVAFXQHFUFHILTT
AVWAFFAWTEVOITDHFHFQAITIXPFHXAFQHEFZQWGFLVWPTOFFA

- Analyze this message using statistics below.
- Ciphertext frequency counts:

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
21	26	6	10	12	51	10	25	10	9	3	10	0	1	15	28	42	0	0	27	4	24	22	28	6	8

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.

Simple Transposition

- Write the message in rows, read out in columns.
- Plaintext: attackxatxdawn
- Ciphertext: ACTWTKXNTXDXAAAX

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

Double Transposition

- Can we do better?
- Plaintext: attackxatxdawn

	Col 1	Col 2	Col 3
Row 1	a	t	t
Row 2	a	c	k
Row 3	x	a	t
Row 4	x	d	a
Row 5	w	n	x

Permute rows and columns

	Col 1	Col 3	Col 2
Row 3	x	t	a
Row 5	w	x	n
Row 1	a	t	t
Row 4	x	a	d
Row 2	a	k	c

- Ciphertext: xtawxnattxadakc
- Key is the matrix size and permutations: `(3,5,1,4,2)` and `(1,3,2)`.
- Often a keyword will indicate the permutation: `STRIPE` $\rightarrow$ `564231`.

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.
  - Guess the keyword!

Cryptanalysis

- Guessing the keyword.

Key Length	No. of permutations	Examples
2	2	AB, BA
3	6	ABC, BAC, CBA
4	24	ABCD, ABDC, ACBD
5	120	ABCDE, ABCED
6	720	ABCDEF, ABDCFE
7	5,040	ABCDEFG, ABDGEF
8	40,320	ABCDEFGH
9	362,880	ABCDEFGHI
10	3,628,800	ABCDEFGHIJ
11	39,916,800	ABCDEFGHIJK
12	479,001,600	ABCDEFGHIJKL

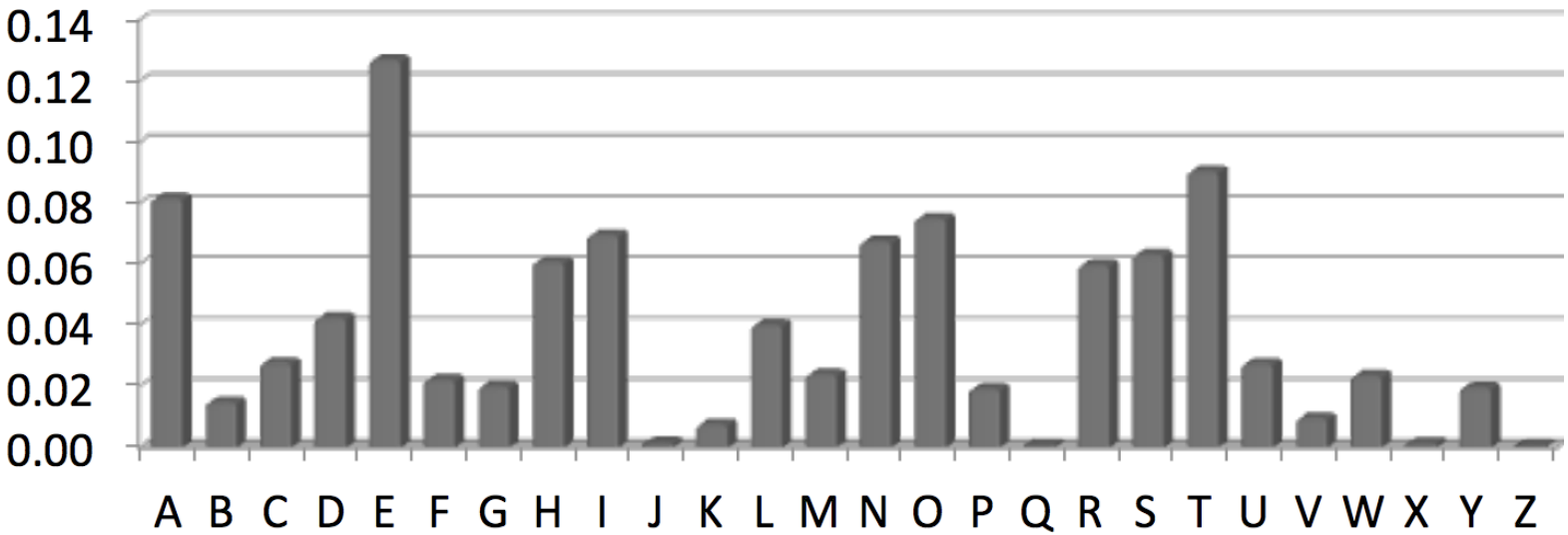
- Can we do better?

Cryptanalysis

- Can we do better?
- Does it make sense to check all (random) keywords?
- Keywords are used to make your life easier, not more difficult!
  - Narrow down the length: unlikely to be very short or very long.
  - Think of possible, meaningful words.
  - “Dictionary attack”.

Cryptanalysis

- But 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?

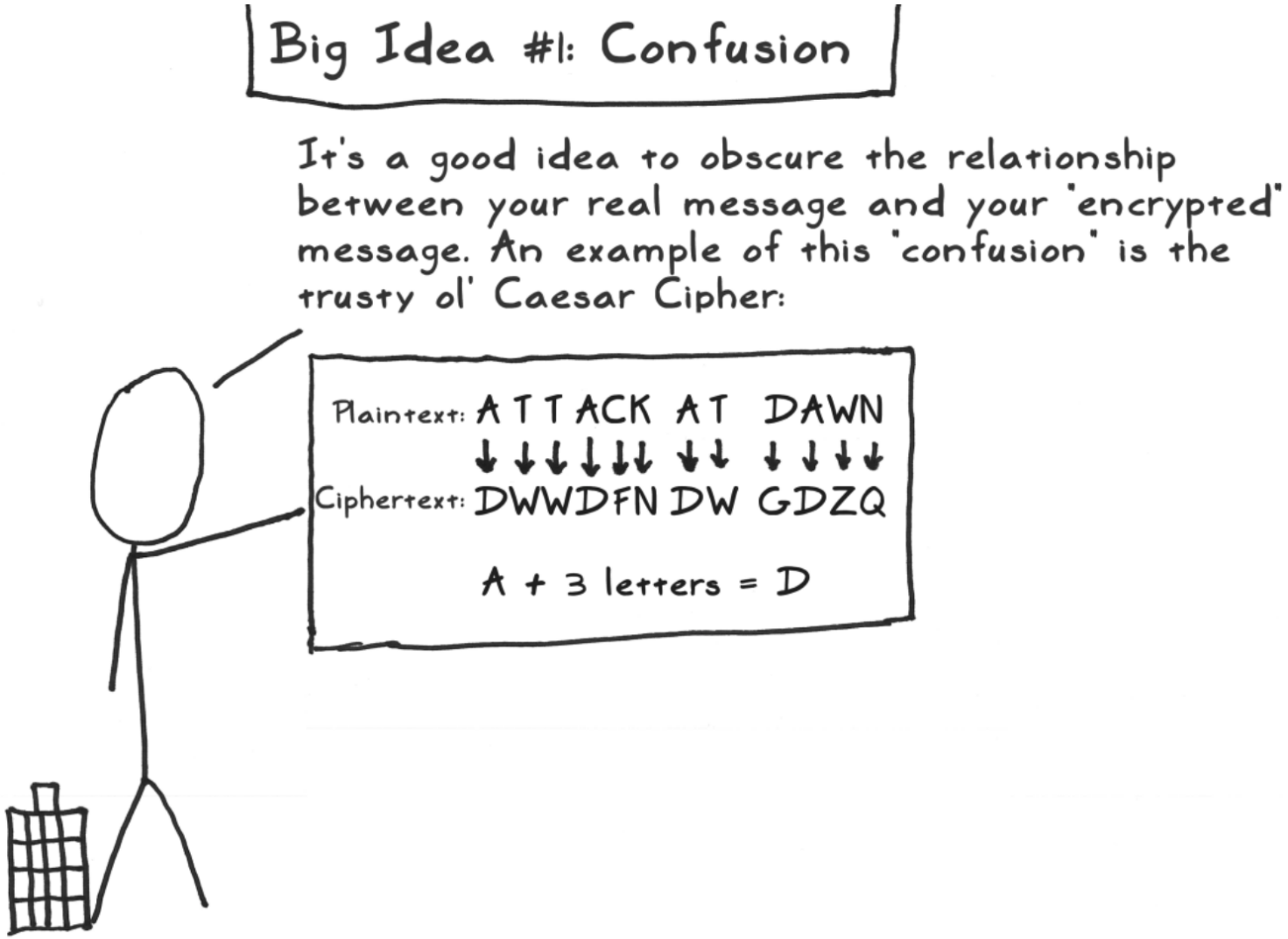
Big Crypto Ideas

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

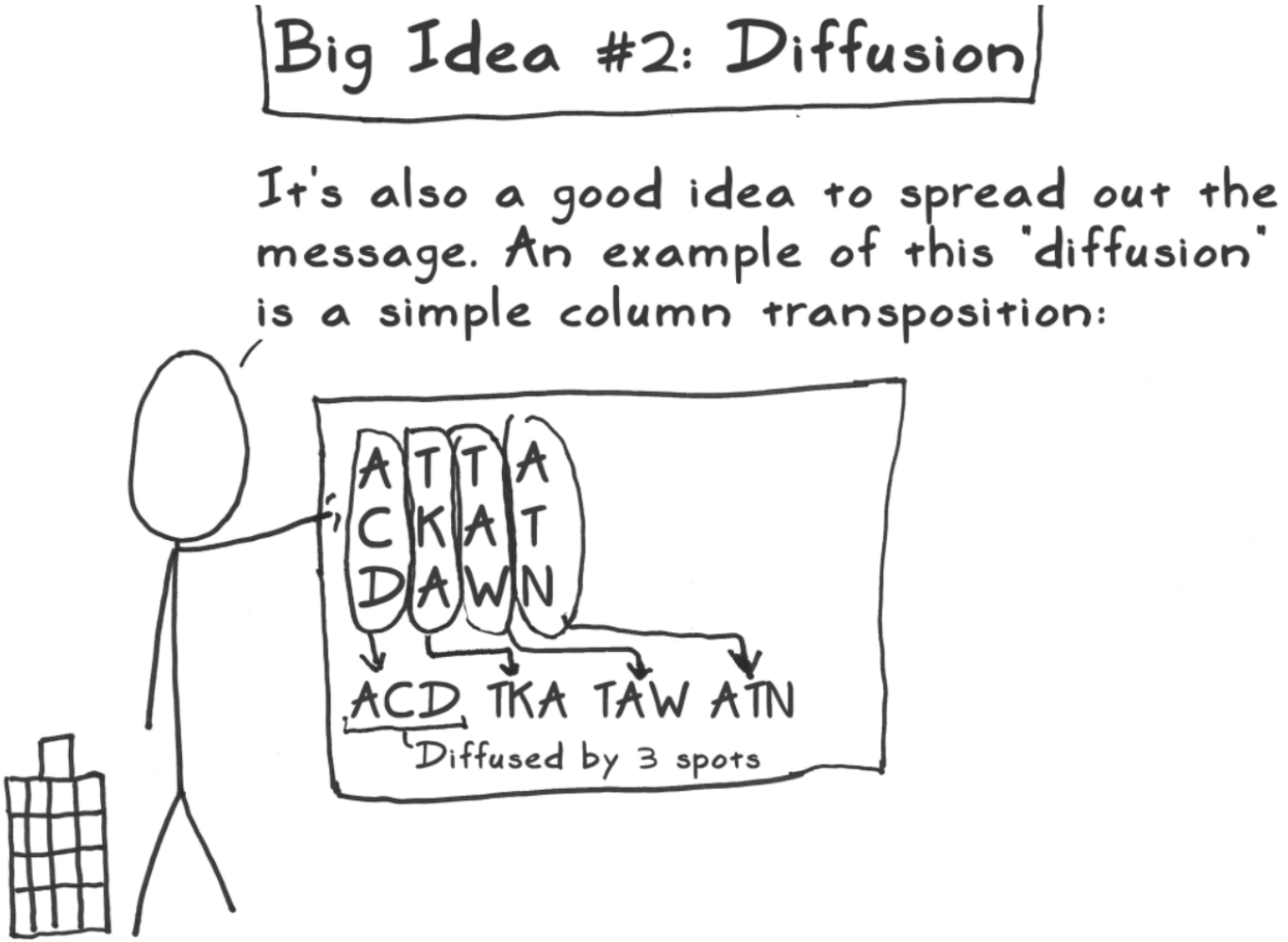
Confusion & Diffusion

- 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*[4].
- DES, AES and many block ciphers are designed using Shannon's idea of confusion and diffusion.

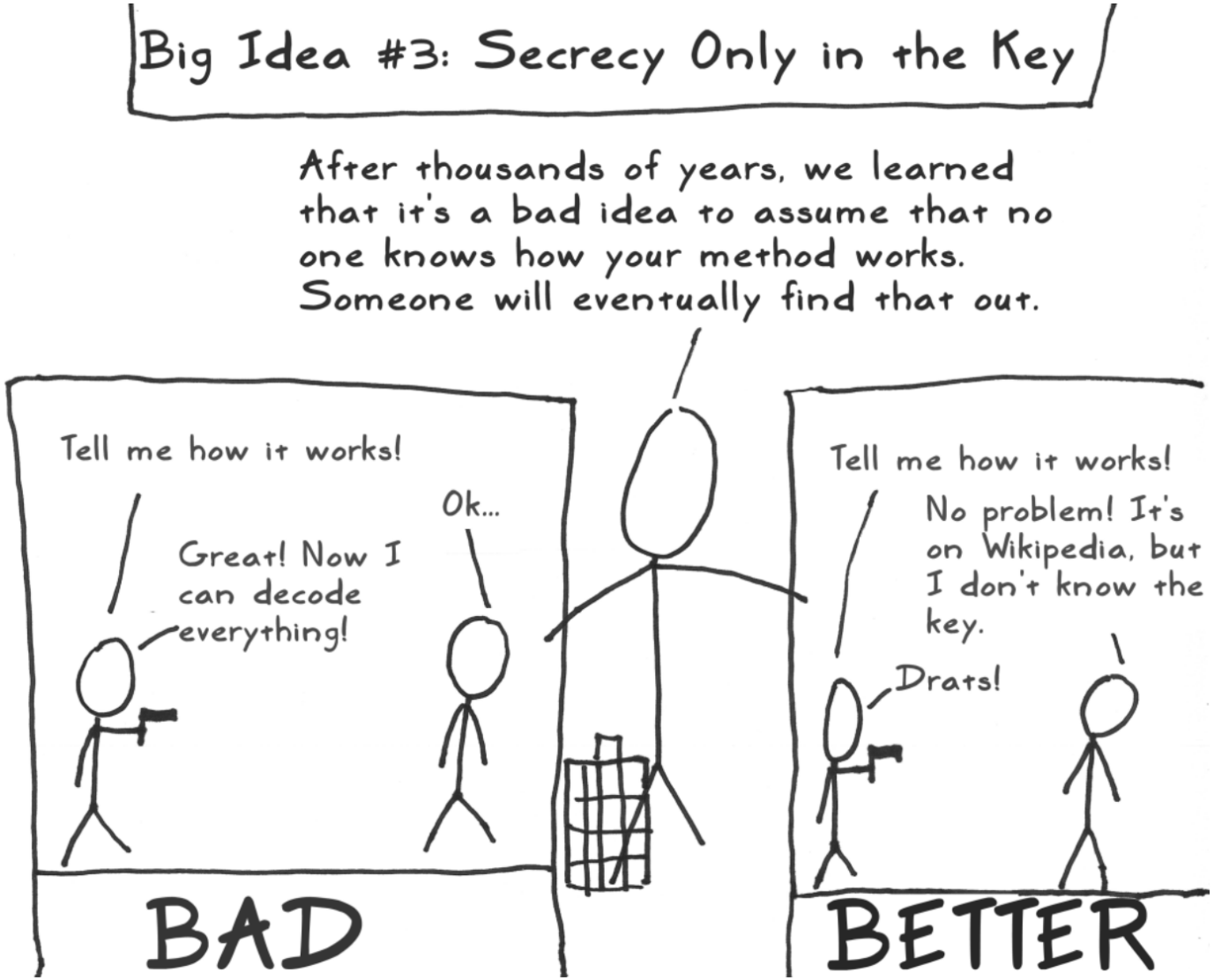
[4] http://netlab.cs.ucla.edu/wiki/files/shannon1949.pdf



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



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



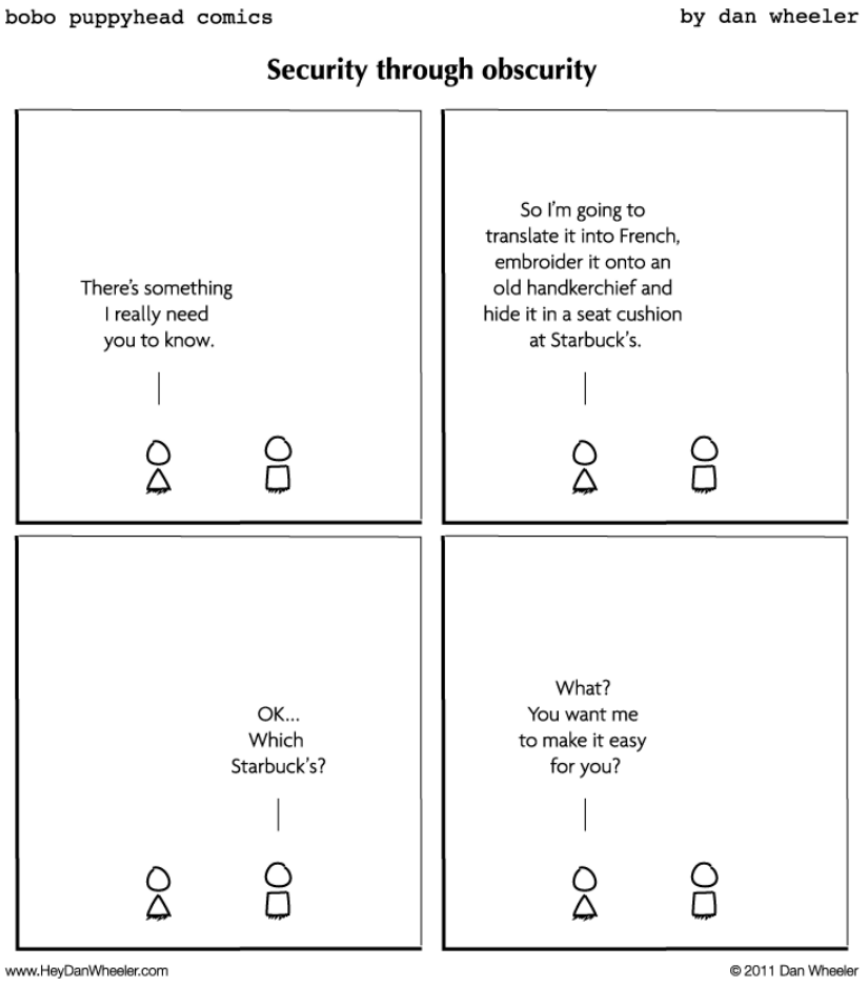
www.moserware.com/2009/09/stick-figure-guide-to-advanced.html

#3: Secrecy in the key

- This is known as Kerckhoffs Principle.
- Basic assumptions:
  - The system is completely known to the attacker.
  - Only the key is secret.
  - That is, crypto algorithms are not secret
- Why do we make such an assumption?
  - Experience has shown that secret algorithms tend to be weak when exposed.
  - Lots of smart people out there!
  - Secret algorithms never remain secret.
  - Better to find weaknesses beforehand.

Security through obscurity is a bad idea![5]

- It just ... is.



[5] Sources: xkcd.com and Dan Wheeler

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!

Avalanche Effect and Evaluation Criteria[6]

- How to evaluate our confusion and diffusion properties?
- Strict avalanche criterion (SAC) states that when a single input bit **i** is inverted, each output bit **j** changes with probability $\frac{1}{2}$, for all **i** and **j**.
- **Translation**: a small change in the plaintext causes a huge change in the ciphertext!
- Bit independence criterion (BIC) states that output bits **j** and **k** should change independently when any single input bit **i** is inverted, for all **i**, **j** and **k**.
- **Translation:** changes in the plaintext with cause random changes in the ciphertext!

[6] https://en.wikipedia.org/wiki/Avalanche_effect

Cryptanalysis: Terminology

- Cryptosystem is secure if best know attack is to try all keys.
  - Exhaustive key search, that is.
- Cryptosystem is insecure if any shortcut attack is known.