Cryptography

Well, a gentle intro to cryptography. Actually, a fairly “hand-wavy” intro to crypto (we’ll discuss why)

Special Thanks: to our friends at the Australian Defense Force Academy for providing the basis for these slides

Definition

- Cryptology is the study of secret writing
- Concerned with developing algorithms which may be used:
  - To conceal the content of some message from all except the sender and recipient (*privacy* or *secrecy*), and/or
  - Verify the correctness of a message to the recipient (*authentication* or *integrity*)
- The basis of many technological solutions to computer and communication security problems

Terminology

- *Cryptography*: The art or science encompassing the principles and methods of transforming an intelligible message into one that is unintelligible, and then retransforming that message back to its original form

- *Plaintext*: The original intelligible message

- *Ciphertext*: The transformed message

- *Cipher*: An algorithm for transforming an intelligible message into one that is unintelligible

Terminology (cont).

- *Key*: Some critical information used by the cipher, known only to the sender & receiver
  - Or perhaps only known to one or the other
- *Encrypt*: The process of converting plaintext to ciphertext using a cipher and a key
- *Decrypt*: The process of converting ciphertext back into plaintext using a cipher and a key

- *Cryptanalysis*: The study of principles and methods of transforming an unintelligible message back into an intelligible message ***without knowledge of the key!***

Still More Terminology…

- *Cryptology*: The field encompassing both cryptography and cryptanalysis
- *Code*: An algorithm for transforming an intelligible message into an unintelligible one using a code-book
  - We'll discuss this only very briefly

Concepts

- Encryption: The mathematical operation mapping plaintext to ciphertext using the specified key:

$$C = E_K(P)$$

- Decryption: The mathematical operation mapping ciphertext to plaintext using the specified key:
  $P = E_K^{-1}(C) = D_K (C)$

- Cryptographic system: The family of transformations from which the cipher function E_K is chosen
  - It is a family of transformations since each key K effectively creates a different transformation

Concepts (cont.)

- *Key*: Is the parameter which selects which individual transformation is used, and is selected from a *keyspace K*
  - How the key is selected is important
    - It's not always uniform!
- More formally we can define the cryptographic system as a single parameter family of invertible transformations

$$E_K \text{ for K in } \mathit{K} \text{ maps P -> C}$$

With unique inverse $P = E_K^{-1}$ for K in *K* maps C -> P

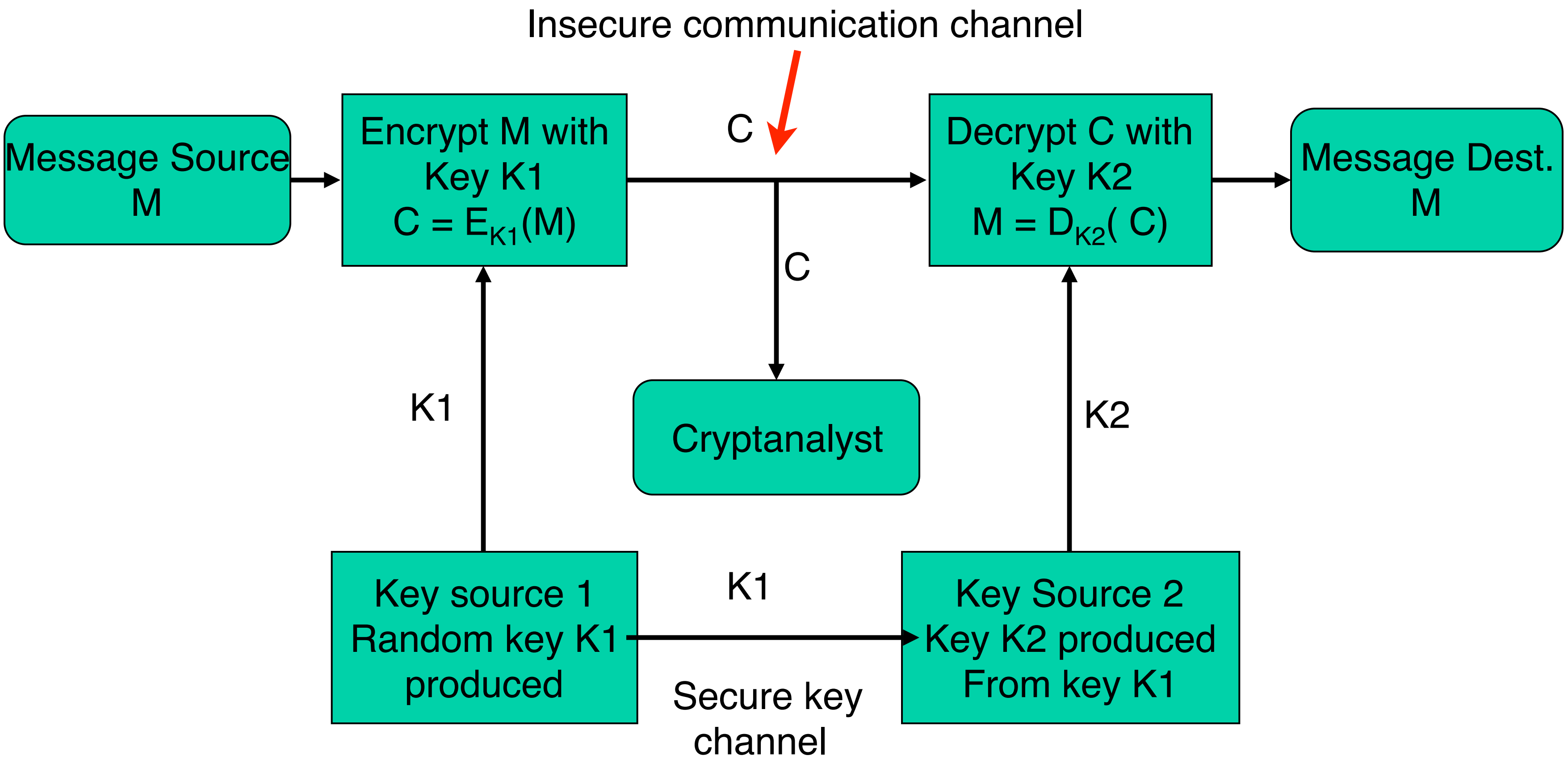
- Usually assume the cryptographic system is public, and only the key is secret information
  - Why?

Rough Classification

- Symmetric-key encryption algorithms
- Public-key encryption algorithms
- Digital signature algorithms
- Hash functions
- Cipher Classes
  - Block ciphers
  - Stream ciphers

We will be discussing each of these (though not all in this slide set)

Symmetric-Key Encryption System


Insecure communication channel
Message Source M
Encrypt M with Key K1 $C = E_{K1}(M)$
C
Decrypt C with Key K2 $M = D_{K2}(C)$
Message Dest. M
C
Cryptanalyst
K1
K2
Key source 1 Random key K1 produced
K1
Key Source 2 Key K2 produced From key K1
Secure key channel

Symmetric-Key Encryption Algorithms

- A Symmetric-key encryption algorithm is one where the sender and the recipient share a common, or closely related, key
  - Managing this key is nontrivial
  - Plus there is the question: how does the key come to be shared?
- Historically, symmetric-key algorithms were developed first
  - They are generally good at efficiently encrypting large amounts of data
    - As of Feb. 2017, an Intel i7 with integrated AES instruction set can encrypt almost 12 GB/s

Types of Cryptanalytic Attacks

- Ciphertext only
  - only know algorithm and some ciphertext
  - use statistical attacks only
    - Probability distributions describing characteristics of plaintext message
  - plus publicly available knowledge
  - must be able to identify when have plaintext
  - Note: This is the most difficult of the classes of attacks we will discuss. For this reason, this is not the attack you want to use to measure the efficacy of an encryption scheme.

Types of Cryptanalytic Attacks

- Known plaintext
  - know (or strongly suspect) some plaintext-ciphertext pairs
  - How?
    - Secret data might not remain secret forever
      - Example: Encrypted message suspected of being contents of official diplomatic statement that is later released
      - Example: If message gives location of an attack (known after attack)
      - Example: Message is text of contract later made public
  - More information gives attacker more leverage. Easier than ciphertext only attack, but still not a good measure of cipher strength

Cryptanalytic Attacks

- Also Partial Plaintext
  - E.g., if message is diplomatic from Russia, expect words such as Moscow, Premier, NATO, etc.
  - Attempt to fill in remaining info using statistical methods

Cryptanalytic Attacks

- Chosen plaintext
  - Can select plaintext and obtain corresponding ciphertext
  - How?
    - Suppose company offers service in which messages are encrypted and transmitted. Attacker trying to read Andrea's confidential message can pay to have the company encrypt any message he (the attacker) wishes
    - Attacker infiltrated senders transmission process so as to be able to cause messages to be encrypted and sent at will
    - Insert records into database and observe changes in statistics after the insertion
  - Especially problematic if attacker knows that ciphertext corresponds to one of a few messages

Chosen Plaintext Attacks in Real World

- Adversary may obtain a device that performs encryptions
- British coercing Germans to send ciphertexts corresponding to locations of mines
- Japanese and attack at Midway

Cryptanalytic Attacks

- Chosen cipher text
  - In addition to what attacker has with a chosen plaintext attack, she can also select ciphertext(s) and obtain corresponding plaintext(s) (with some limits)
  - How?
    - Any time the attacker has a "decryption oracle"
    - Such as a tamper-resistant cryptographic smart card system!
    - Many crypto systems require this level of security
    - And some of these don't have it!
    - Also called "lunchtime attack", as reference to attacker having access to decryption device while employee out to lunch
- Adaptive-chosen cipher text attack.

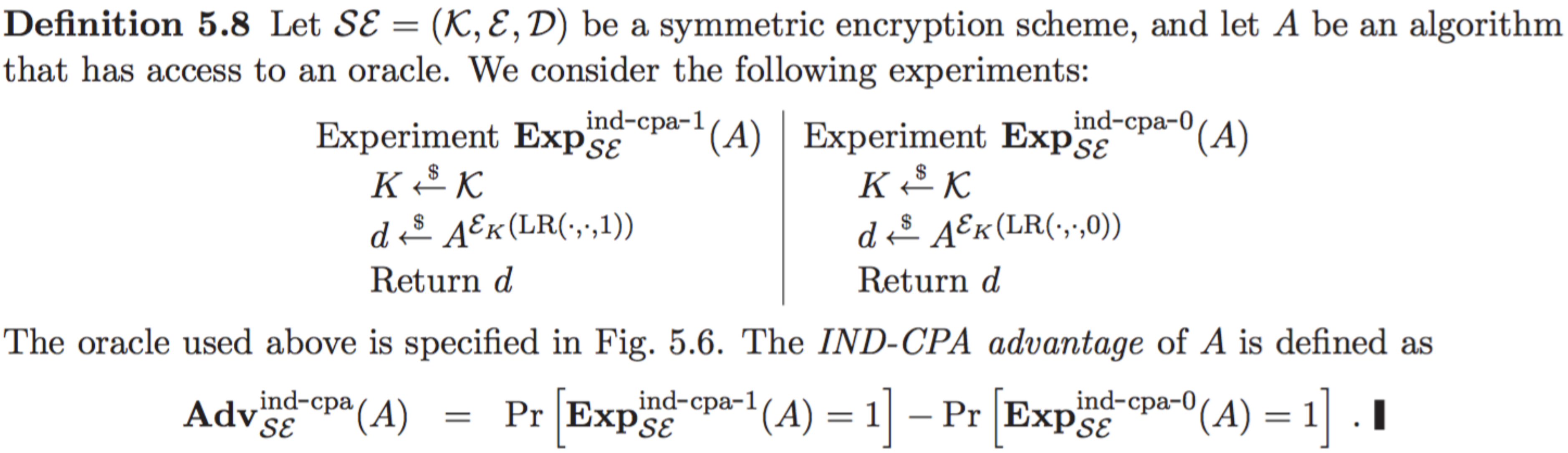
Chosen Ciphertext Attack in Real World

- Midway example: US forces could have sent cipher texts to Japanese and monitored their behavior
- Adversary sends encrypted messages to banks and then monitors the behavior of the bank (how does it respond, etc).
- One common example: encryption algorithm often used as part of a higher level protocol. E.g., used as part of an authentication protocol, where one party sends a ciphertext to another, who decrypts it.

Cryptanalytic Attacks

- A good cipher *must* resist all four attacks!
  - Typically, chosen ciphertext attack is the one that we measure against, since this gives the adversary the most weapons.

More Rigorous Definitions: Security against Chosen Plaintext Attack



Definition 5.8 Let $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be a symmetric encryption scheme, and let A be an algorithm that has access to an oracle. We consider the following experiments:

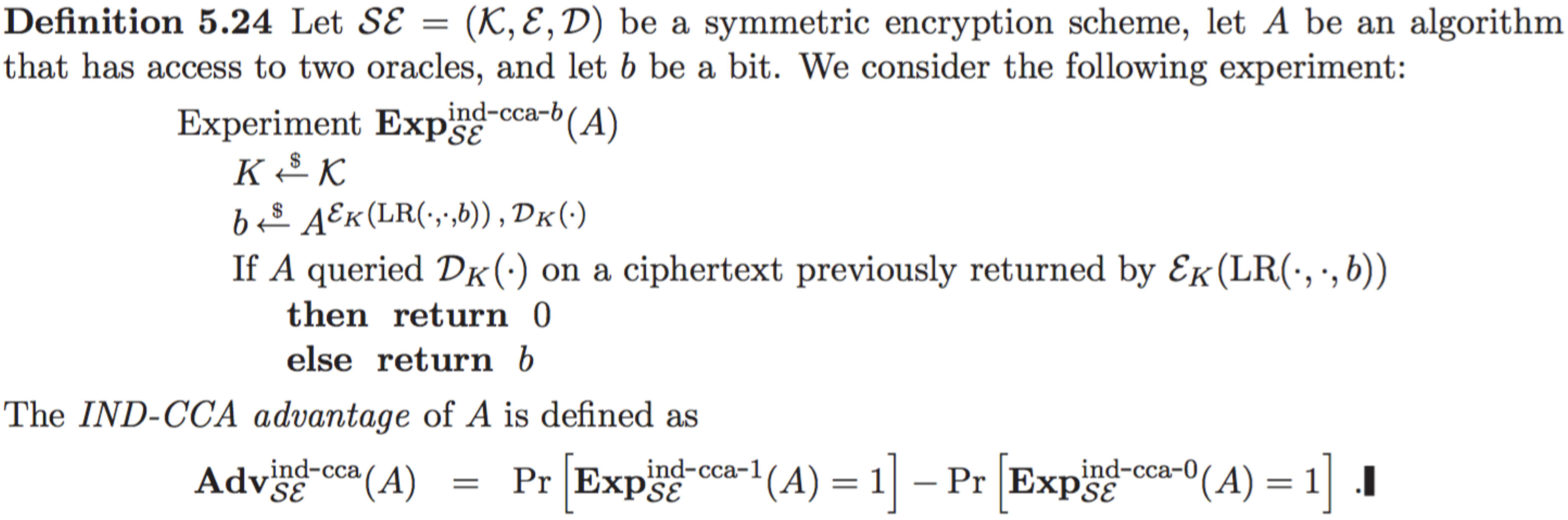
```
Experiment $\mathbf{Exp}_{\mathcal{SE}}^{\text{ind-cpa-1}}(A)$
    $K \xleftarrow{\$} \mathcal{K}$
    $d \xleftarrow{\$} A^{\mathcal{E}_K(\mathrm{LR}(\cdot,\cdot,1))}$
    Return $d$
```

```
Experiment $\mathbf{Exp}_{\mathcal{SE}}^{\text{ind-cpa-0}}(A)$
    $K \xleftarrow{\$} \mathcal{K}$
    $d \xleftarrow{\$} A^{\mathcal{E}_K(\mathrm{LR}(\cdot,\cdot,0))}$
    Return $d$
```

The oracle used above is specified in Fig. 5.6. The *IND-CPA advantage* of A is defined as

$$\mathbf{Adv}_{\mathcal{SE}}^{\text{ind-cpa}}(A) \;=\; \Pr\left[\mathbf{Exp}_{\mathcal{SE}}^{\text{ind-cpa-1}}(A) = 1\right] - \Pr\left[\mathbf{Exp}_{\mathcal{SE}}^{\text{ind-cpa-0}}(A) = 1\right] . \blacksquare$$

More Rigorous Definitions: Security against Chosen Ciphertext Attack



Definition 5.24 Let $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be a symmetric encryption scheme, let A be an algorithm that has access to two oracles, and let b be a bit. We consider the following experiment:

```
Experiment $\mathbf{Exp}_{\mathcal{SE}}^{\text{ind-cca-}b}(A)$
  $K \stackrel{\$}{\leftarrow} \mathcal{K}$
  $b \stackrel{\$}{\leftarrow} A^{\mathcal{E}_K(\text{LR}(\cdot,\cdot,b)),\, \mathcal{D}_K(\cdot)}$
  If $A$ queried $\mathcal{D}_K(\cdot)$ on a ciphertext previously returned by $\mathcal{E}_K(\text{LR}(\cdot,\cdot,b))$
     then return 0
     else return $b$
```

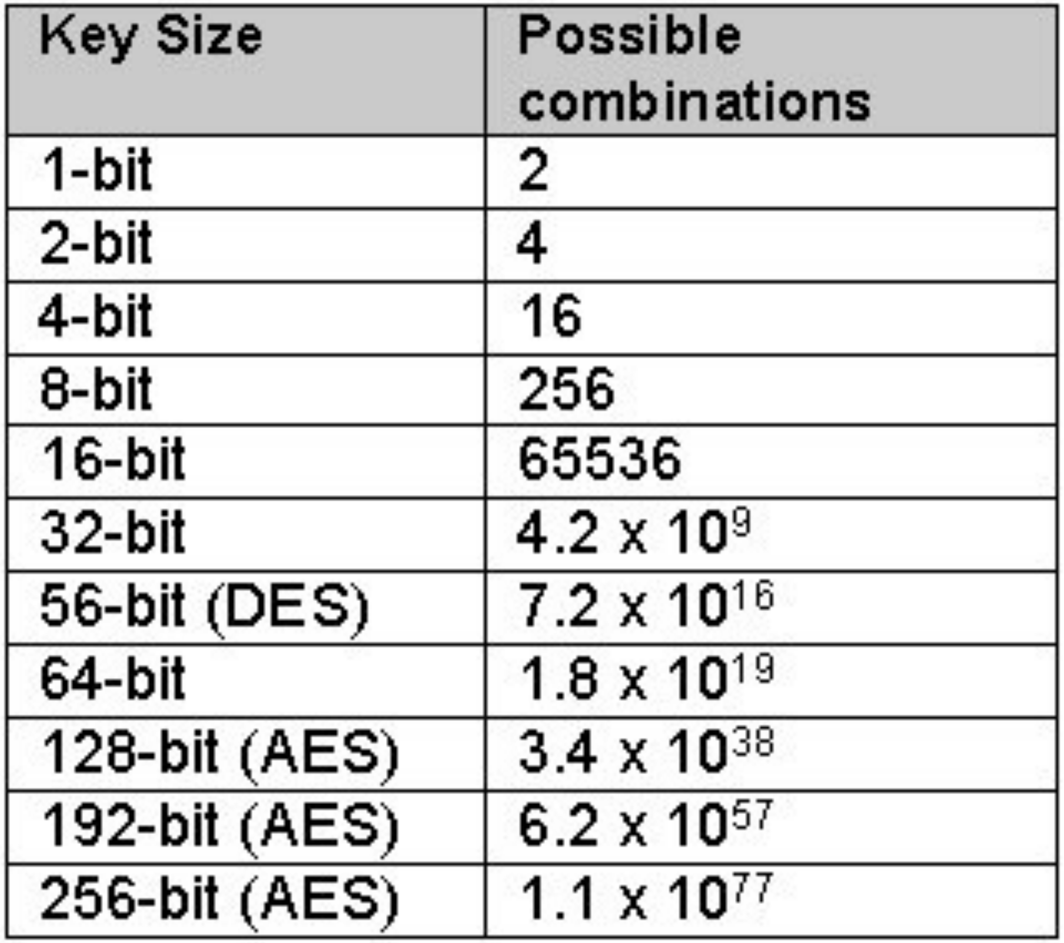
The *IND-CCA advantage* of A is defined as

$$\mathbf{Adv}_{\mathcal{SE}}^{\text{ind-cca}}(A) \;=\; \Pr\left[\mathbf{Exp}_{\mathcal{SE}}^{\text{ind-cca-1}}(A) = 1\right] - \Pr\left[\mathbf{Exp}_{\mathcal{SE}}^{\text{ind-cca-0}}(A) = 1\right] . \blacksquare$$

Exhaustive Key Search

- Always theoretically possible to simply try every key
- Most basic attack, directly proportional to key size
- *Assumes attacker can recognize when plaintext is found!!*

Exhaustive Key Search



Key Size	Possible combinations
1-bit	2
2-bit	4
4-bit	16
8-bit	256
16-bit	65536
32-bit	4.2 x 10^9
56-bit (DES)	7.2 x 10^{16}
64-bit	1.8 x 10^{19}
128-bit (AES)	3.4 x 10^{38}
192-bit (AES)	6.2 x 10^{57}
256-bit (AES)	1.1 x 10^{77}

Exhaustive Key Search

- Fastest Supercomputer (Wikipedia): As per June 2012, IBM Sequoia
  - 16.31 Petaflops = 16.31 x 10^{15} FLOPS
- Number of FLOPS required per key check
  - Optimistically estimated at 1000
- Number of key checks per second
  - 16.31 x 10^{15} / 1000 = 16.31 x 10^{12}
- Number of seconds in a year
  - 31,536,000
- Number of years to crack 128-bit AES
  - (3.4 x $10^{38)}$ / [(16.31 x 10^{12}) x 31536000] = 6.61 x 10^{17}

Exhaustive Key Search

- Fastest Supercomputer (Wikipedia): As per June 2018, Summit (in the U.S.)
  - 122.3 Petaflops = 122.3 x 10^{15} FLOPS
- Number of FLOPS required per key check
  - Optimistically estimated at 1000
- Number of key checks per second
  - 122.3 x 10^{15} / 1000 = 122.3 x 10^{12}
- Number of seconds in a year
  - 31,536,000
- Expected number of years to crack 128-bit AES = 4.411 x 10^{16}

Unconditional and Computational Security

- Unconditional security: No matter how much computer power is available, the cipher cannot be broken since the ciphertext provides insufficient information to uniquely determine the corresponding plaintext
  - Probabilistic definition: basically, having ciphertext gives no info about plaintext

- Computational security: Given limited computing resources (e.g., time needed for calculations is greater than age of universe), the cipher cannot be broken

Classic Encryption Techniques

- Two basic components in classical ciphers: substitution and transposition

- *Substitution ciphers* - letters replaced by other letters

- *Transposition ciphers* – same letters, but arranged in a different order

- Several such ciphers may be concatenated together to form a *product cipher*

The Caeser Cipher

- 2000 years ago Julius Caesar used a simple substitution cipher, now known as the *Caesar cipher*
  - First attested use in military affairs (e.g., Gallic Wars)
- Concept: replace each letter of the alphabet with another letter that is k letters after original letter
- Example: replace each letter by 3rd letter after

L FDPH L VDZ L FRQTXHUHG

I CAME I SAW I CONQUERED

The Caeser Cipher

- Can describe this mapping (or translation alphabet) as:

  Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ Cipher: DEFGHIJKLMNOPQRSTUVWXYZABC

General Caesar Cipher

- Can use any shift from 1 to 25
  - I.e. replace each letter of message by a letter a fixed distance away
- Specify *key letter* as the letter a plaintext A maps to
  - E.g. a key letter of F means A maps to F, B to G, ... Y to D, Z to E, I.e. shift letters by 5 places
- Hence have 26 (25 useful) ciphers
  - Hence breaking this is easy. Just try all 25 keys one by one.

Mathematics

- If we assign the letters of the alphabet the numbers from 0 to 25, then the Caesar cipher can be expressed mathematically as follows:

For a fixed key k, and for each plaintext letter p, substitute the ciphertext letter C given by

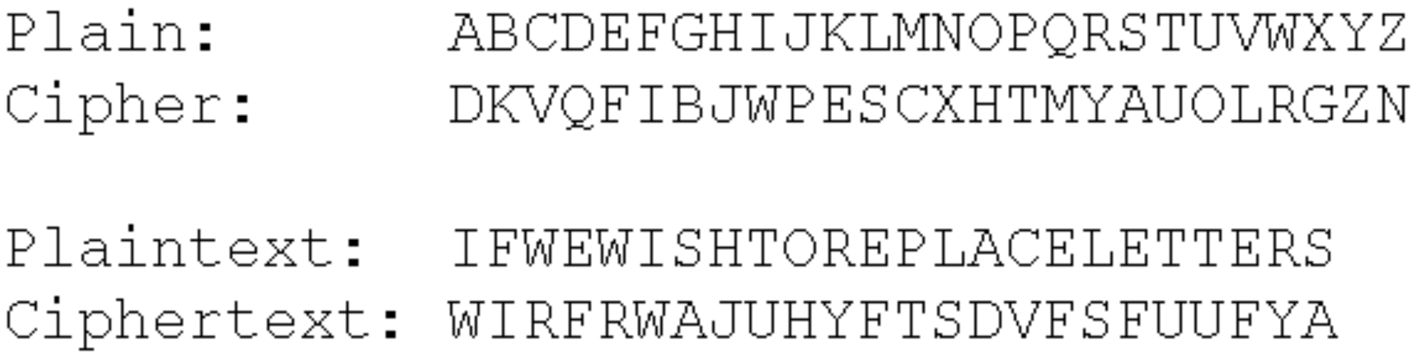
$$C = (p + k) \bmod(26)$$

Decryption is equally simple:

$$p = (C - k) \bmod (26)$$

Mixed Monoalphabetic Cipher

- Rather than just shifting the alphabet, could shuffle (jumble) the letters arbitrarily
- Each plaintext letter maps to a different random ciphertext letter, or even to 26 arbitrary symbols
- Key is 26 letters long



```
Plain:       ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher:      DKVQFIBJWPESCXHTMYAUOLRGZN

Plaintext:   IFWEWISHTOREPLACELETTERS
Ciphertext:  WIRFRWAJUHYFTSDVFSFUUFYA
```

Security of Mixed Monoalphabetic Cipher

- With a key of length 26, now have a total of 26! ~ 4 x 10^{26} keys
    - A computer capable of testing a key every ns would take more than 12.5 billion years to test them all.
    - On average, expect to take more than 6 billion years to find the key.
- With so many keys, might think this is secure…but you'd be wrong

Security of Mixed Monoalphabetic Cipher

- Variations of the monoalphabetic substitution cipher were used in government and military affairs for many centuries into the middle ages

- The method of breaking it, *frequency analysis* was discovered by Arabic scientists

- All monoalphabetic ciphers are susceptible to this type of analysis

Language Redundancy and Cryptanalysis

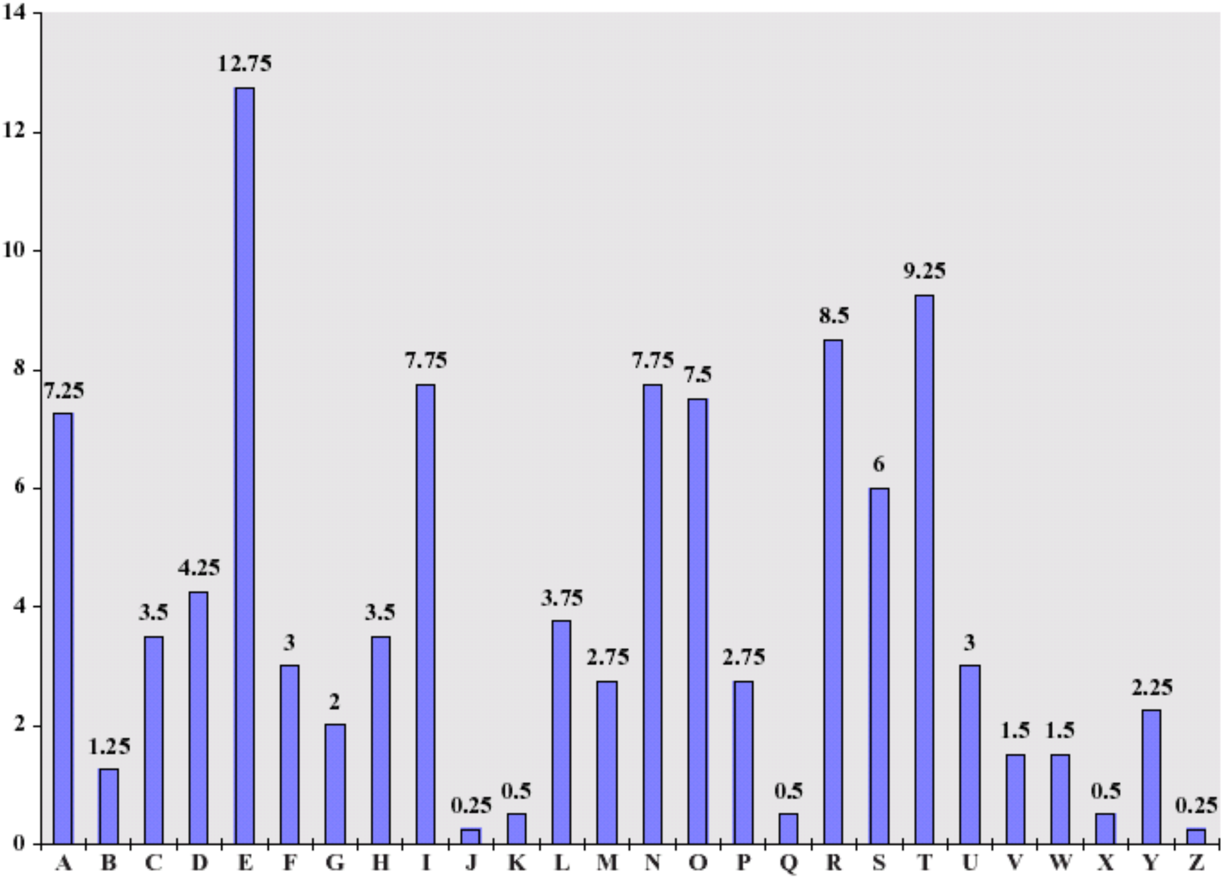
- Human languages are redundant
- Letters in a given language occur with different frequencies.
  - Ex. In English, letter e occurs about 12.75% of time, while letter z occurs only 0.25% of time.
- In English the letters e is by far the most common letter

Language Redundancy and Cryptanalysis

- t,r,n,i,o,a,s occur fairly often, the others are relatively rare
- w,b,v,k,x,q,j,z occur least often
- So, calculate frequencies of letters occurring in ciphertext and use this as a guide to guess at the letters. This greatly reduces the key space that needs to be searched.

Language Redundancy and Cryptanalysis

- Tables of single, double, and triple letter frequencies are available



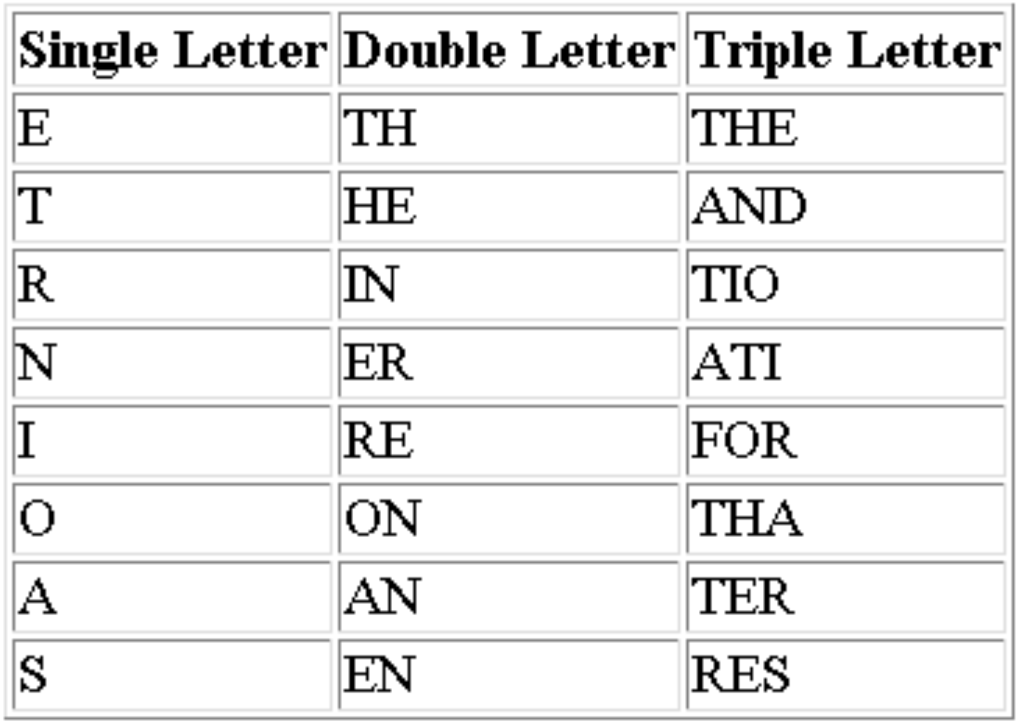
Other Languages

- Natural languages all have varying letter frequencies
- Languages have different numbers of letters (cf. Norwegian)
- Can take sample text and count letter frequencies
- Seberry (1st Ed) text, Appendix A has counts for 20 languages. Hits most European & Japanese & Malay

Performing Frequency Analysis

- Calculate letter frequencies for ciphertext being analyzed
- Compare counts/plots against known values
- In particular look for common peaks and troughs
  - Peaks at: A-E-I spaced triple, NO pair, RST triple with U shape
  - Troughs at: JK, X-Z
- Key concept - monoalphabetic substitution does not change relative letter frequencies

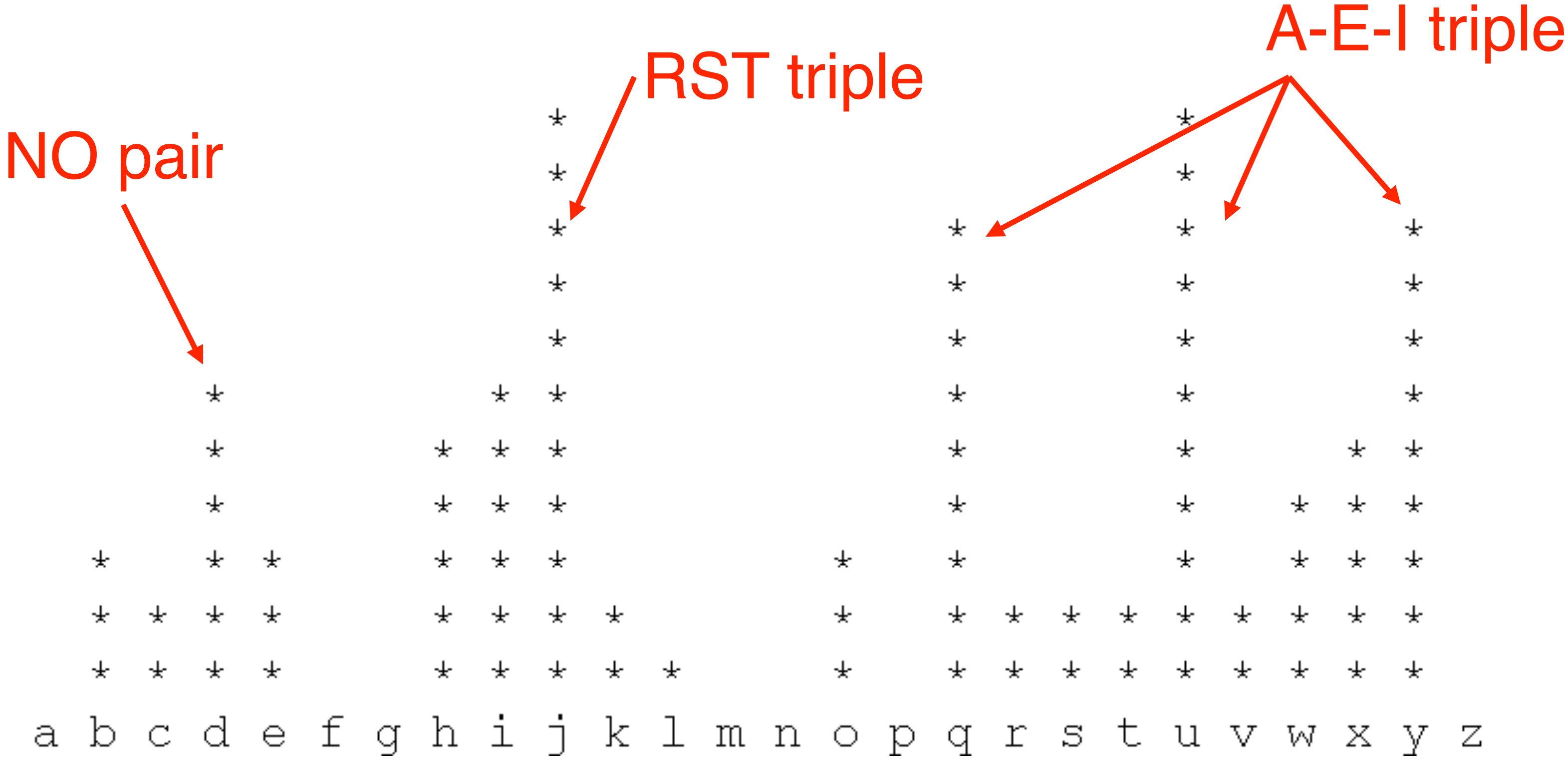
Table of Common English Single, Double and Triple Letters



Single Letter	Double Letter	Triple Letter
E	TH	THE
T	HE	AND
R	IN	TIO
N	ER	ATI
I	RE	FOR
O	ON	THA
A	AN	TER
S	EN	RES

Example with Caesar Cipher

- given "JXU WHUQJUIJ TYISELUHO EV COWUDUHQJYED YI JXQJ Q XKCQD UYDW SQD QBJUH XYI BYVU RO QBJUHYDW XYI QJJYJKTUI"



NO pair

RST triple

A-E-I triple

Polyalphabetic Ciphers

- Might guess that one approach to improving security is to use multiple cipher alphabets, hence the name polyalphabetic ciphers
- Makes cryptanalysis harder since have more alphabets to guess and because flattens frequency distribution
- Use a key to select which alphabet is used for each letter of the message
  - ith letter of key specifies ith alphabet to use
- Use each alphabet in turn
- Repeat from start after end of key is reached

But…

- Cryptanalysts have methods for determining the key length
  - E.g., if two identical sequences of plaintext occur at a distance that is an integer multiple of the key length, then their ciphertext will be identical
  - Ex: key: `DECEPTIVEDECEPTIVEDECEPTIVE`

    Plaintext: `WEAREDISCOVEREDSAVEYOURSELF`

    Ciphertext: `ZICVTWQNGRZGVTWAVZHCQYGLMGJ`
- Once you have key length, cracking this is just cracking multiple monoalphabetic ciphers

Book Cipher

- If key length is the issue with polyalphabetic cipher, at limit want as many alphabets as letters in message (but how to transfer such a key if it's truly random?)
- Book cipher: create key as long as a message by using words from a book to specify the translation alphabets
- Key used is then the book and page and paragraph to start from
- British used this some in WWII (called them poem codes)
  - Big problem

Problems with Book Cipher

- Same language characteristics are used by the key as the message
  - i.e., a key of 'E' will be used more often than a 'T' etc, hence an 'E' encrypted with a key of 'E' occurs with probability $(0.1275)^2$ = 0.01663, about twice as often as a 'T' encrypted with a key of 'T'
- Have to use larger frequency table, but they exist
- Given sufficient ciphertext this can be broken

- BUT, if a truly random key as long as the message is used, the cipher is **provably** unbreakable
  - Called a *One-Time Pad*

One-Time Pad

- A true solution: Choose a random key as long as the message itself
  - This reveals nothing statistically about the plaintext message. This lack of information about plaintext means that a one-time pad is unbreakable.

One-Time Pad

- Practical considerations
  - Sender and receiver must be in possession of, and protect, the random key. If the receiver loses the key, they will have no way to reconstruct the plaintext.
  - Can only use a given key once, since if used even as few as two times, cryptanalysis reduces to frequency analysis on digraphs
  - Rarely used in practice (often no point in using it, since key is as long as the message)
    - But once both parties have key, can transmit many messages (until sum of lengths reach length of key)
  - Implementation issues have also led to one-time pad systems being broken

Transposition Ciphers

- Also known as **permutation** ciphers
- Core idea: hide the message by rearranging the letter order without altering the actual letters used
- Can recognize these since have the same frequency distribution as the original text
- Very Simple Example: Mirror Cipher (write message backwards). Obviously not very secure
  - But what about mirror image in Russian?!

Cracking Transposition Ciphers

- Cracking transposition ciphers involves educated guessing with much trial and error
- BUT, there is software that will do a lot of this stuff for you (and it’s out there and freely available)
- Bottom line, neither substitution nor transposition ciphers are secure (with the exception, of course, of a well-implemented one-time pad).

Increasing Cipher Security

- Ciphers based on just substitutions or transpositions are not secure
- Several ciphers in succession might seem to make cryptanalysis more difficult, but:
  - two substitutions are really only one more complex substitution
  - two transpositions are really only one more complex transposition
- A substitution followed by a transposition, however, makes a new much harder cipher
  - We call these *product ciphers*

Steganography

- an alternative to encryption
- hides existence of message
  - using only a subset of letters/words in a longer message marked in some way
  - using invisible ink
  - hiding in LSB in graphic image or sound file
- has drawbacks
  - high overhead to hide relatively few info bits
  - If adversary realizes you’re using steganography, you’re usually sunk