SJSU SAN JOSÉ STATE UNIVERSITY

Lesson 1 – Classical Ciphers

Yan Chen

CS166 Fall 2024

- yan.chen01@sjsu.edu
  - Add [CS166] at the beginning of the subject line
- Everything on Canvas, check it regularly
  - Assignments, announcements, lecture notes, etc.
- In person sessions will be recorded
  - Lecture part only, excluding homework hint
- Office hour on Zoom (Link on Canvas)
  - Hours: Regular: T/Th 13:50 – 14:50, first come, first serve
  - Or appointment (Link on Canvas)

- At least 130 pts in total
  - Assignments (7 * 3 = 21 pts)
  - Midterms (2 * 3 = 6 pts)
  - (Mandatory) Final (100 pts) = max(final, sum of midterms)
  - Others (3+ pts)
- Grading Scale
  - Raw points, not percentages

Grade	Pts	Grade	Pts	Grade	Pts
A	≥ 93.00	**B minus**	80.00 to 82.99	**D plus**	66.00 to 69.99
A minus	90.00 to 92.99	**C plus**	76.00 to 79.99	**D**	63.00 to 65.99
B plus	86.00 to 89.99	**C**	73.00 to 75.99	**D minus**	60.00 to 62.99
B	83.00 to 85.99	**C minus**	70.00 to 72.99	**F**	≤ 59.99

- Class protocols
  - Passwords only given in class (in person & in recording)
  - No sharing course materials
  - No late homework question via Email
  - NO cheating!
- Effective communication
- Important Dates
  - Sep. 17, Tuesday: Last day to drop without a W grade
  - Dec. 17, Tuesday: Final Exam 14:45 – 17:00 PT

    (can pick earlier dates)

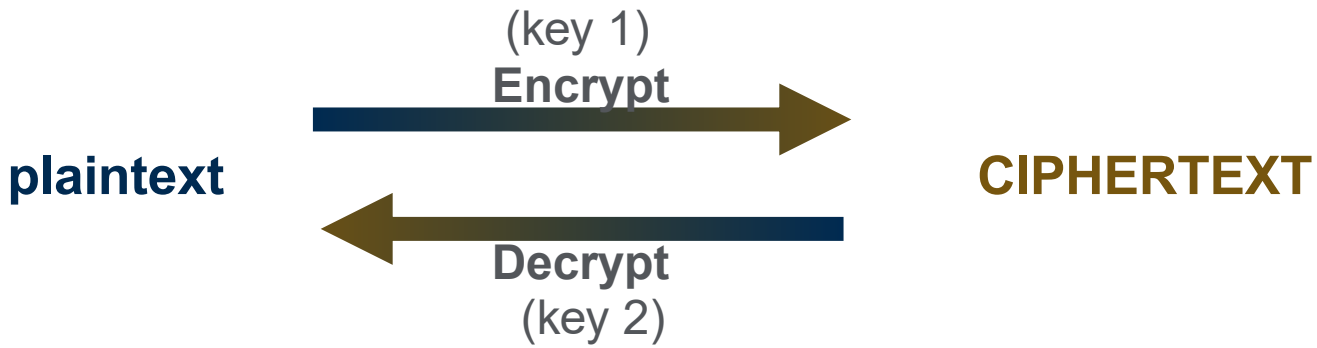
- The Cast of Characters
  - Alice (customer) and Bob (system) are the good guys
  - Trudy is the attacker
- CIA triad: primary focus
  - Confidentiality: prevent unauthorized reading of information
  - Integrity: detect unauthorized writing of information
  - Availability: data is available in a timely manner when needed
- 4 topics
  - Crypto, software, access control, protocols
- Think like Trudy, but NOT act like Trudy!

- Alice and Bob want to communicate (exchange information) secretly
  - Again, Alice & Bob not necessarily human
  - They don't want other people know the information exchanged
- When the distance between Alice & Bob is large, they have to communicate via a channel (media)
  - Old school: Pigeons, mails, phone
  - Digital era: network, Internet
- But these channels are not secure...
  - That's why we need "Cryptology"...

- Cryptography: making “secret codes” (secret message)
  - Note that here “code” does not mean “program”
- Cryptanalysis: breaking “secret codes”
  - Alice & Bob want prevent Trudy from doing that!
- Cryptology: making and breaking “secret codes”
  - Cryptology = Cryptography + Cryptanalysis
- Crypto: a synonym for any of the above and more!
- Cipher system (cryptosystem): a suite of algorithms needed to implement a particular security service

- Cipher system elements
  - Plaintext: the original data (in lowercase)

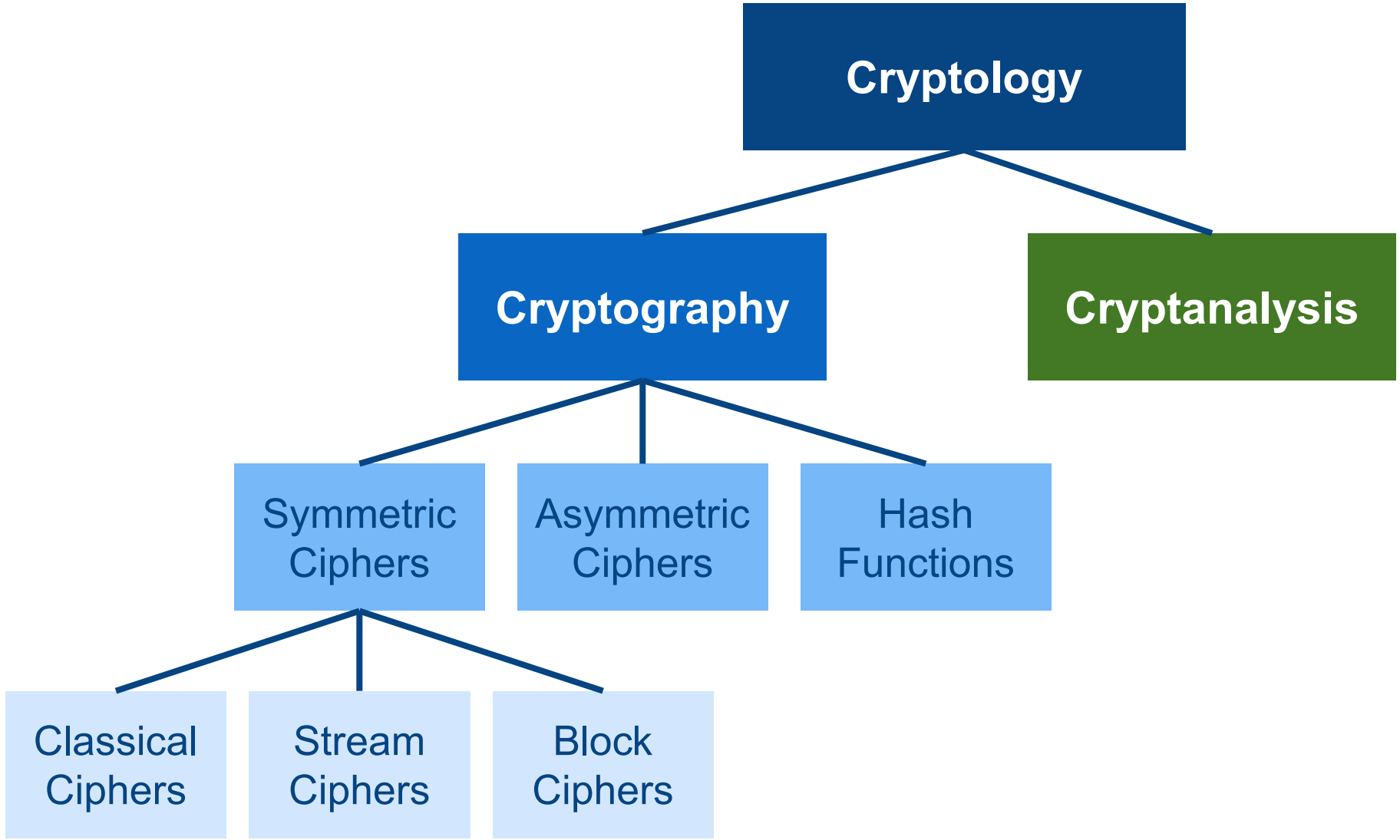
    (Data can be any form, such as text, audio, video, ...)
  - Encryption: convert plaintext to ciphertext
  - Ciphertext: the result of encryption (in uppercase)
  - Decryption: convert ciphertext back to plaintext
  - Key(s): string(s) for configuring the cipher system



- Key used for encryption and decryption can be different
  - Symmetric cipher system: same key (symmetric key) are used for both encryption and decryption
  - Asymmetric cipher system: different keys are used (public key for encrypt, private key for decrypt)
- Typically, a cipher system consists 3 algorithms
  - One algorithm for key generation
  - One algorithm for encryption
  - One algorithm for decryption

- Kerckhoffs' principle: the strength of a cryptosystem depends ONLY on the key
  - Trudy knows the system (algorithm & ciphertext)
  - Trudy only doesn't know the key (and of course, the plaintext)
- Because experience has shown that ...
  - Secret algorithms tend to be weak
  - Secret algorithms never remain secret
  - Better to find weaknesses beforehand
- Cryptographers will not use a cryptosystem until it has been approved by many cryptographers over time

Classification of Cryptology



Motivation

- Classical ciphers are dead!
  - We don't use it anymore!
- Then why need to talk about classical ciphers?
  - Some of them represent the features of modern ciphers
  - Roughly speaking, modern ciphers are the enhanced version of the classical ciphers by combining those features
- So, we need to analyze the features of the following classical ciphers, and learn why they are dead
  - Simple substitution
  - Double transposition, one-time pad, codebook

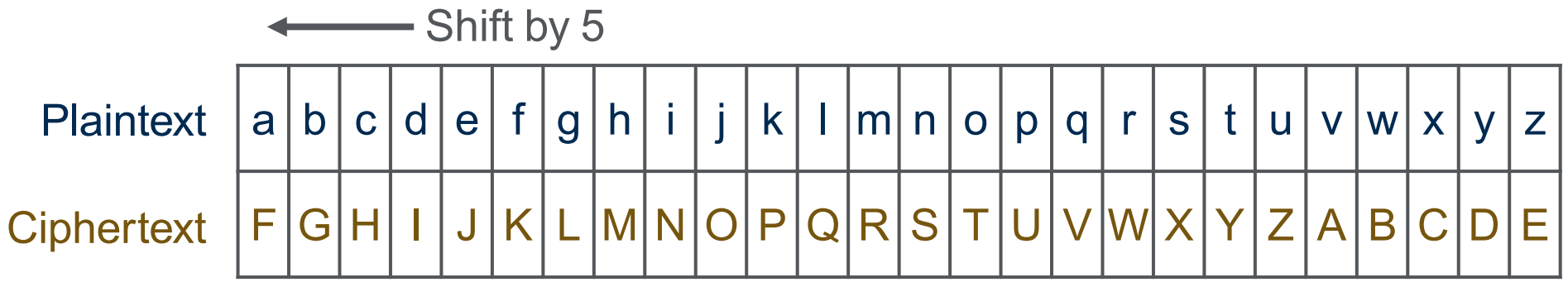
- Simple substitution: each letter is substituted with another (one-to-one mapping)
- Simplest: Caesar's cipher – shift left by alphabet by 3
  - Used 2,000 years ago, named after Roman emperor Caesar

Shift by 3

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

- Examples
  - Encrypt "hello" → "KHOOR"
  - Decrypt "ZRUOG" → "world"

- A little enhancement: parameterize the key
  - Instead of hard-code the key = 3
  - i.e., shift by k for some k ∈ {1, 2, …, 25}
  - And let's call it "parameterized Caesar cipher"
- Example: k = 5



  - Encrypt "hello" → "MJQQT"
  - Decrypt "BTWQI" → "world"

- Cryptanalysis of parameterized Caesar
  - i.e., how can Trudy find the key?
  - The simplest way: checking all possible values of k
  - So how many values does Trudy need to try?

    Worst case: 25 attempts; Average: 13 attempts (~half)
- Exhaustive key search (brute-force attack): check the whole keyspace
  - Keyspace: the set of all possible values of the key

    e.g. $k \in \{1, 2, \ldots, 25\}$ is the keyspace of parameterized Caesar
  - This attack is always available for Trudy! (💡 why?)

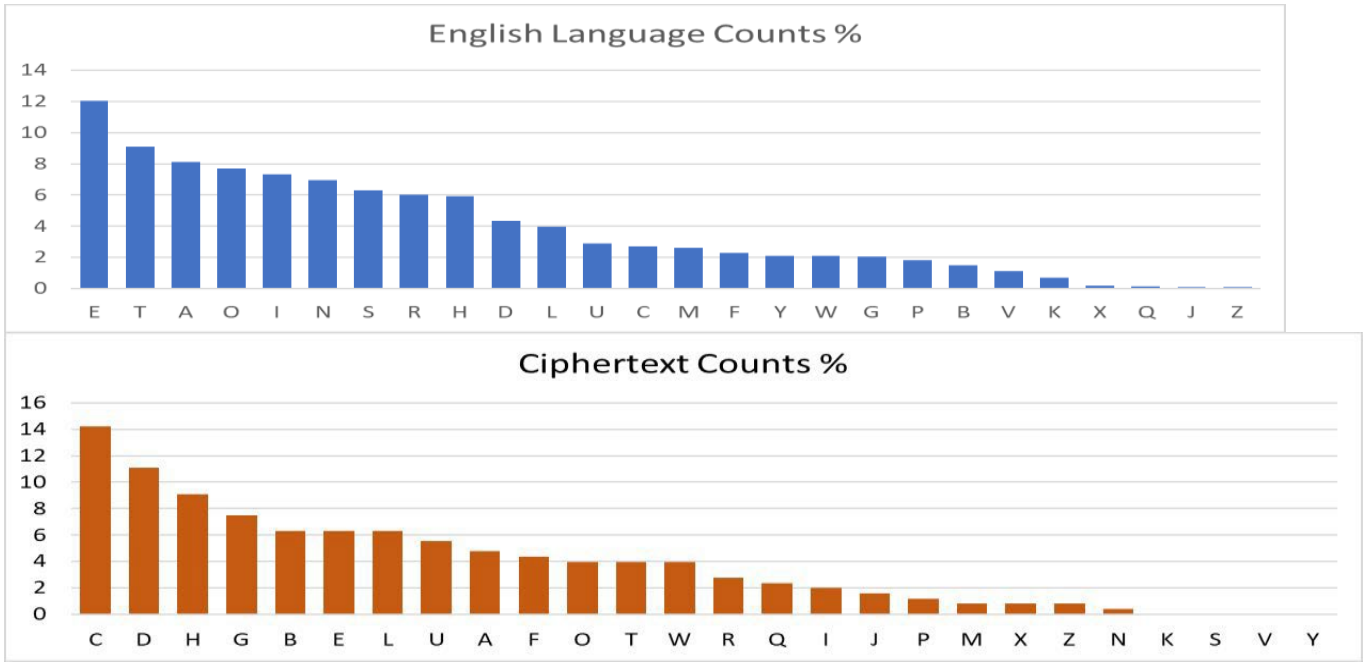
- In general, the key to a simple substitution cipher can be any permutation of letters
  - Instead of simple shift
  - Also known as “character cipher” or “monoalphabetic cipher”
- Example

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

  - Encrypt “hello”
  - Decrypt “UTHWA”

- What is the size of the keyspace?
  - i.e., how many possible keys (permutations)?
  - $26! \approx 2^{88}$ keys!
  - i.e., if use exhaustive key search, Trudy need to try 2^{88} times in the worst case; and on average, need $2^{88} / 2 = 2^{87}$ times
- How many years does Trudy need if she used a computer that can check 2^{20} ($\approx$ 1 million) keys/second?
  - On average: $2^{87} / 2^{20} = 2^{67}$ sec $\approx 4.7 * 10^{12}$ years!
- The larger keyspace makes general simple substitution "stronger" than parameterized Caesar, but...

- Any smarter way (shortcut) to break simple substitution?
  - Yes, use linguistic knowledge...English letter frequency!
  - Also called “statistical attack”
  - For example ...



  - Need large enough ciphertext though

- A large keyspace is necessary but not sufficient to ensure security of a cipher
  - ➢ Only ensures that exhaustive key search is infeasible
  - ➢ But there can be a shortcut!
- So simple substitution is proven insecure!
  - ➢ Vulnerable to statistical attacks!
- Definition of “secure” for a cipher system...
  - ➢ A cipher system is secure if best know attack is to try all keys
  - ➢ A cipher system is insecure if any shortcut attack is known
  - 💡 Under this definition, is parameterized Caesar secure?

- Motivation: hide the statistics of the letters (“diffusion”)
  - Prevent statistical attacks
- (Simplified) double transposition cipher
  - Put plaintext into a matrix (1 letter / cell)
  - Permutate the rows and columns
  - Key is the matrix size and permutations
- Example plaintext: wanttogetana
  - key: 3 * 4 matrix,

    (1, 3, 4, 2) and (2, 1, 3)
  - Ciphertext: awnegtnaatto

	1	2	3
1	w	a	n
2	t	t	o
3	g	e	t
4	a	n	a

→

	2	1	3
1	a	w	n
3	e	g	t
4	n	a	a
2	t	t	o

- To decrypt, just undo the permutation
- Pros: hide the statistic
  - In previous example, a is substituted to w, t, or o, NOT to a single letter as in simple substitution
- Cons: cipher does not disguise the letters
  - Just shuffled the order of the letters
  - If can find several "words", possible to break
- Idea is employed by modern block ciphers
  - Deal with a "block" of text

	1	2	3
1	w	a	n
2	t	t	o
3	g	e	t
4	a	n	a

	2	1	3
1	a	w	n
3	e	g	t
4	n	a	a
2	t	t	o

- Motivation: hide relationship between plaintext and ciphertext ("confusion")
- One-time pad: key only used once, and use XOR
- Recap for XOR (exclusive OR)
  - Commutative: $p \oplus q = q \oplus p$
  - Associative: $(p \oplus q) \oplus r = q \oplus (p \oplus r)$
  - Identity: $p \oplus 0 = p$
  - Self-Inverse: $p \oplus p = 0$

p	q	p ⊕ q
0	0	0
0	1	1
1	0	1
1	1	0

- Encryption/decryption based on the property of XOR: if $p \oplus q = r$, then $p = q \oplus r$ (💡 how to prove?)

- Procedures of using one-time pad
  - First, encode plaintext to binary (encoding rule is public)
  - Then, randomly generate a binary string in the same size of the encoded plaintext as the key
  - To encrypt: CIPHER = plain $\oplus$ key, then decode to text
  - To decrypt: plain = CIPHER $\oplus$ key, then decode to text
- Example
  - Encoding rule: a = 00, n = 01, t = 10, w = 11

plaintext	w	a	n	t	a	n	a
(encoded) p	11	00	01	10	00	01	00
key	01	10	11	00	11	10	01
(encoded) C	10	10	10	10	11	11	01
CIPHERTEXT	T	T	T	T	W	W	N

- Pros: provably secure
  - Ciphertext gives NO useful info about plaintext
  - All plaintexts are equally likely
  - But only when used properly –use random & one-time key!
  - What will happen if the key is reused?
- Cons: not practical
  - Recall: key size is same as the length of plaintext
  - If we have a secure channel to send the key...why not directly send the plaintext itself?
- One-time pad is developed to modern stream ciphers

- Codebook: dictionary-like book filled with “codewords”
  - Words (plaintext) and corresponding codewords (ciphertext)
  - The code book itself is the key
- Example
  - Codebook (key):

word	codeword
a	10928
to	31287
get	09165
want	82096
...	...

  - Plaintext: want to get a
  - Ciphertext: 82096 31287 09165 10928

- The security of this cipher system depends on the physical security of the codebook
  - If the amount of ciphertexts is big enough, the statistical attack is possible (but harder than simple substitution)
- So, codebook usually use with “additive”
  - Additive: book of “random” numbers
  - Key is the codebook + position in additive book (which gives a Message Indicator MI)
  - For each word, new cipher = old cipher + MI
- Modern block ciphers are codebooks!

- Stream cipher
  - A5/1
  - RC4

- Terminologies
  - Cryptography, Cryptanalysis, Cryptology, Cryptology, Crypto
- Cipher system
  - Plaintext, ciphertext, encryption, decryption, key, keyspace
  - Symmetric vs. Asymmetric (public/private)
  - Exhaustive key search
  - Secure vs. insecure
  - Confusion vs. diffusion
- Kerckhoffs' principle
- Classical ciphers
  - Caesar Cipher, parameterize cipher, simple substitution
  - Double transposition, one-time pad, codebook

- Integrity or confidentiality, which one (and why) is more important from the perspective of...
  - The bank
  - The bank's customers
- Edgar Allan Poe's 1843 short story, "The Gold Bug," features a cryptanalytic attack
  - What type of cipher is broken and how?
  - What happens as a result of this cryptanalytic success?
- Given that the Caesar's cipher was used, find the plaintext that corresponds to the following ciphertext

VSRQJHEREVTXDUHSDQWV

- Suppose we keep the spaces and punctuations as they are when we use simple substitution. Break the following message.
  - DAHUFJOU HUC OCECBDA REDTGDWR TBFPACQ LG REFIE HF PC ETWFQTACHC, HUCBC LG D GTCWLDA WDGC HUDH WDE PC GFAZCX LE ALECDB HLQC. D GJTCBLEWBCDGLEO REDTGDWR LG GLQLADB HF HUC OCECBDA REDTGDWR CNWCTH HUDH, IUCE HUC ICLOUHG DBC DBBDEOCX MBFQ ACDGH HF OBCDHCGH, CDWU ICLOUH LG OBCDHCB HUDE GJQ FM DAA TBCZLFJG ICLOUHG.
  - What if we remove the spaces and punctuations?

    Is it harder to break or easier?

References

- Stamp, Mark and Low, Richard M., “Applied Cryptanalysis: breaking ciphers in the real world,” John Wiley & Sons, Inc., New Jersey, USA, 2007
- Stallings, William, “Cryptography and Network Security, Principles and Practice, 6th ed.,” Pearson, USA, 2014
- Paar, Christof, “Understanding Cryptography,” Faller, Berlin, Germany, 2010
- Stamp, Mark, “Information Security, Principles and Practice, 2nd ed.,” Wiley, New Jersey, USA, 2011