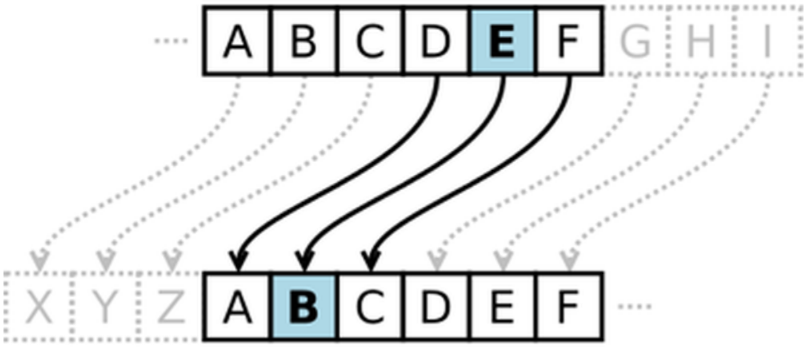
Crack the Code: Breaking a Caesar Cipher

Summary

AREAS OF SCIENCE

Cybersecurity

Computer Science



DIFFICULTY

TIME REQUIRED

Short (2-5 days)

PREREQUISITES

Experience with Python or another programming language of your choice.

MATERIAL AVAILABILITY

This project requires a computer.

SAFETY

No issues

CREDITS

Ben Finio, PhD, Science Buddies

Abstract

When you hear the word "encryption," you might think about modern computers and things like email and online bank accounts. But did you know that encryption has been around for thousands of years? In this project you will learn about the Caesar cipher, a simple type of encryption that replaces each letter of the alphabet with another letter, and demonstrate how a modern computer can crack this ancient code in just a few seconds.

Have you ever wanted to send a secret message to a friend? What if someone, like a parent or a teacher, intercepts the message and reads it? In order to make sure that only your friend could read the message, even if it was intercepted, first you would need to encrypt it. **Encryption** is the process of encoding a message so only the intended recipient can read it. Encryption is used to protect many of our daily online activities, like emails and credit card transactions, from unauthorized access.

Modern encryption algorithms are very complicated and (ideally) difficult to break. However, encryption has been around for thousands of years—long before computers existed. Leaders throughout history have used various types of encryption to send messages to allied countries and military leaders during wartime. One famous example is the **Caesar cipher**, used by Julius Caesar in ancient Rome. The Caesar cipher is an example of a **substitution cipher**, where each letter of the alphabet (in English, 26 letters) is replaced by another letter of the alphabet. This is done by "shifting" the entire alphabet by a certain number of spaces. This number is called the **key**. For example, here is a shift of 3 (note how the alphabet "wraps around" from the end):

Original alphabet:	`ABCDEFGHIJKLMNOPQRSTUVWXYZ`
Shifted alphabet:	`DEFGHIJKLMNOPQRSTUVWXYZABC`

To encode a message, each letter in the original message (called the **plaintext**) is replaced with the letter directly below it in the shifted alphabet (A becomes D, B becomes E, and so on). The result is called the **ciphertext**. Here is a plaintext message encrypted using a shift of 3:

Plaintext:	`THIS IS A SECRET MESSAGE`
Ciphertext:	`WKLV LV D VHFUHW PHVVDJH`

In order to share secret messages, you and your friend need to agree on a key in advance. Then, you can use the key to encrypt messages, and your friend can use the same key (shifting the alphabet in the opposite direction) to decrypt them. Anyone who intercepts the messages will be unable to read them if they do not know the key.

But, what if a very determined person wants to crack your code? How could they do it? One major weakness of the Caesar cipher is that it is vulnerable to a **brute-force attack**, an attack that tries all possible keys to decrypt a message. Since there are only 25 possible keys in English (using a key of 26 gets you back to the original alphabet), for very short encrypted messages it would not take you long to manually try all the keys. For example, here is a short encrypted message (note that this simple version of the Caesar cipher only changes letters; punctuation remains unchanged).

QOB MCI QFOQY HVS QCRS?

The message is still gibberish, so we know that 1 is not the key (assuming the original message was actually in English!). Can you try to decrypt the message using the other 24 possible keys? Keep trying different keys until you get a sentence that makes sense in English. How long does it take you to do it by hand?

Another method that can be used to crack a Caesar cipher (or any other type of substitution cipher) is **frequency analysis**. Frequency analysis is based on the fact that certain letters appear with different frequencies in English writing—for example, E usually occurs the most often, followed by T and A; whereas Q and Z appear the least often (Figure 1).

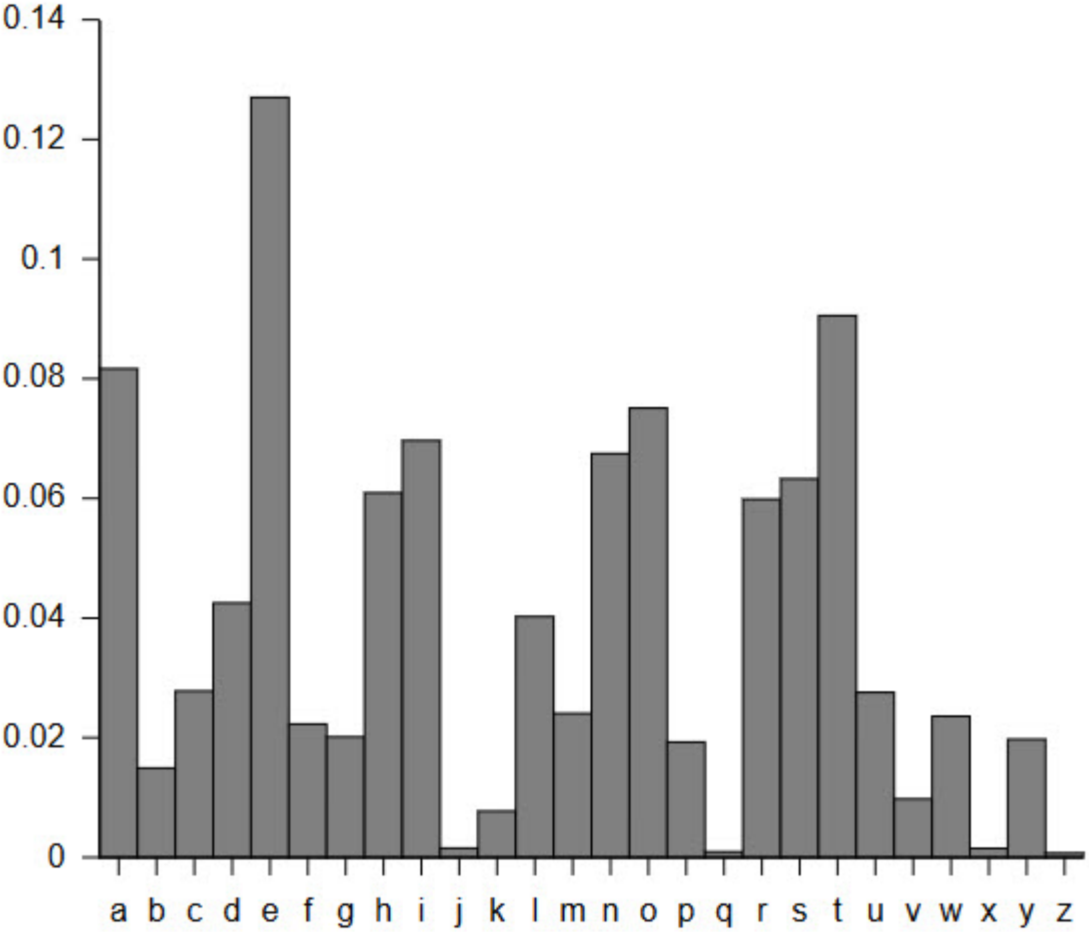


Figure 1. Letter frequencies in the English language.

For example, look at this encrypted text:

L PZ AOL TVZA MYLXBLUA SLAALY PU AOPZ ZLUALUJL decrypt this message using a key of 7 (L becomes E, M becomes F, and so on)?

Doing a brute-force attack or frequency analysis by hand can be easy for very short messages, but can become time-consuming for entire paragraphs or pages of text. This is where writing a computer program to do the work for you comes in handy. In the procedure of this project, you will write your own programs that can first encrypt plaintext using a Caesar cipher, and then attempt to decrypt the text using both a brute-force attack and frequency analysis.

Terms and Concepts

- Encryption
- Caesar cipher
- Substitution cipher
- Key
- Plaintext
- Ciphertext
- Brute-force attack
- Frequency analysis
- Modulo

Questions

- What is a substitution cipher?
- How does a Caesar cipher work?
- How is a Caesar cipher vulnerable to attacks? Why does this make it a poor choice for modern encryption?

Bibliography

- Learn Cryptography (n.d.). *Caesar Cipher*. Retrieved July 11, 2017.
- Rodriguez-Clark, D. (n.d.). *Frequency Analysis: Breaking the Code*. Crypto Corner. Retrieved July 11, 2017.
- Shaw, Z. (n.d.). *Learn Python the Hard Way*. Retrieved July 11, 2017.

- Lab notebook

Experimental Procedure

Cybersecurity Project Warning

Cybersecurity projects can be fun, but they can also get you in trouble if you are not careful. Make sure you follow these rules when doing a cybersecurity project:

- Do not attack any individual, computer, system, or network without consent from the individual (or person who owns the computer). For example, do not try to guess someone's email password and log into their account unless you get their permission first, or try to hack into a website without permission from the owner of the website.
- Even if you have consent to perform an attack, the attack should be for learning purposes only, and you should help the individual or organization fix any problems you find (this is known as "white hat" hacking). For example, if you are able to guess someone's password, you should tell them they need to pick a stronger password (and help them learn how). Do not read their emails, change any of their account settings, look at private information or files like pictures, or tell anyone else their password.
- If your project involves human subjects, even if you have their consent, you may still need approval from your science fair or an Institutional Review Board (similar to the rules for psychology or medical experiments). See this page for more information.
- Do not pretend to be a different person, company, or other organization online. This includes pretending to be someone else on a social media site, setting up fake websites designed to look like real websites from reputable companies, or sending "phishing" or other emails designed to look like they were sent by someone else. (A controlled experiment where only study participants have access to examples of such websites or emails would be OK.)
- Do not use data that was illegally obtained (for example, contact information stolen from a company's employee database), even if it was stolen by someone else and already posted online.
- Do not publicly post sensitive personal information, even if it was obtained with consent. For example, if your project involves accessing people's contact information (legally), do not post someone's name and address in the "Results" section of your science fair display board. You should destroy any such information (by shredding paper or deleting files) when you are done with your project.
- Do not install or run any malicious software (viruses, malware, spyware, trojans, etc.) on a

administrator before you start.

If you are doing this project in Python, you might want to make sure you know how to use the following features of the language before you start (or equivalent features in a program of your choice).

- Lists
- Strings
- IF statements
- FOR loops
- The **modulo** operator (%)

If you get stuck when writing your program, an online search for something general (like "python if statement") or specific (like "how to read a string from a text file in python") will typically give helpful results.

1) On your computer, write a sentence or short paragraph (or copy one from this page) and save it as a text file.

2) Write a program that:

a. Reads a plaintext string from a text file.

b. Encrypts the string using a Caesar cipher with a randomly generated key. You can make your program only change the letters A-Z and leave other characters (numbers, punctuation, spaces) unchanged.

c. Saves the ciphertext to a new text file.

3) Write a program to perform a brute-force attack on the ciphertext. Refer to the background section if you need a reminder about how a brute-force attack works. The program should:

a. Load the encrypted string from the text file.

b. Try all 25 possible keys to decrypt the ciphertext, saving each result in a new string.

c. Look at all 25 resulting strings. Most of them should be gibberish. Do any of them make sense? Can you figure out which one was the correct key?

4) Write a program to perform frequency analysis on the ciphertext. Refer to the background section if

c. Use this information to calculate the key (assuming the most common letter corresponds to the letter E in plaintext).

d. Decrypt the text using the key you calculated. Does the resulting plaintext make sense? If not, what do you think went wrong? (hint: be careful with frequency analysis, E might not be the most common letter in individual sentences or short paragraphs)

5) Test your program. Search online for text that has already been encrypted with a Caesar cipher (so you cannot "cheat" by already knowing the answer) and try using your program to decrypt it. You can also test your program on the following blocks of text. Which approach works better for each message, brute force or frequency analysis?

```
Example 1: BNMFQZSTKZSHNMR! XNT GZUD BQZBJDC SGD BNCD!
Example 2: UZU PFL KYZEB KYZJ GIFAVTK NRJ WLE? TYVTB FLK KYV
           CVRIE DFIV JVTKZFE KF CVRIE RSFLK TRIVVIJ ZE TPSVIJVTLIZKP
```

Ask an Expert

Do you have specific questions about your science project? Our team of volunteer scientists can help. Our Experts won't do the work for you, but they will make suggestions, offer guidance, and help you troubleshoot.

Post a Question

Variations

- Can you expand your Caesar cipher program so it also encrypts other characters (letters, punctuation, spaces)?
- The Caesar cipher is just one type of substitution cipher. There are many other types of substitution ciphers, including more complicated types that are designed to defeat frequency analysis. Can you write a program to encrypt and decrypt messages using a different type of cipher?
- This project requires that you check the results of your decryption program manually to see if the decryption worked. This can still be time consuming if you need to decrypt many separate messages. Can you automate this process? (hint: do a web search for "python check if a word is English")
- Frequency analysis is less reliable for short blocks of text where E might not be the most common letter. Examine various chunks of text (for example, taken from your favorite website or book) of

might have no idea how it was encrypted or even what language it was written in. Can you write a program that attempts to decrypt messages using multiple types of substitution cipher? What about a program that works on messages written in Spanish or another language?

- Find a friend to work with. Write your own encryption algorithm and challenge your friend to crack it, and vice versa.

- Share your program with someone else and use it to encrypt and decrypt messages that you send each other (for example, do the encryption on your computer, send the encrypted text via email, and the recipient can decrypt on their computer). Do you think it is secure to keep using the same key forever? Can you come up with a system to change the key, for example based on the date?

Careers

If you like this project, you might enjoy exploring these related careers:

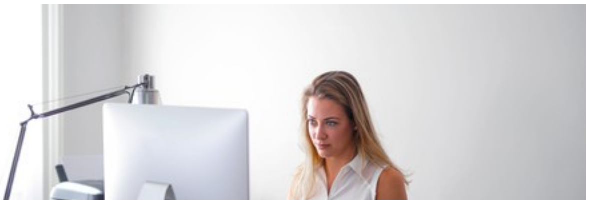
Information Security Analyst *In Demand!*

Career Profile

Have you ever seen a story on the news about how a company or government agency was "hacked" and people's personal information, like names, addresses, or credit card numbers, was stolen? It is an information security analyst's job to prevent that from happening. Organizations hire information security analysts to analyze possible threats against their computer systems, which can range from malicious hackers trying to steal data to careless employees who accidentally forget to log out of a… Read more

Penetration Tester *In Demand!*

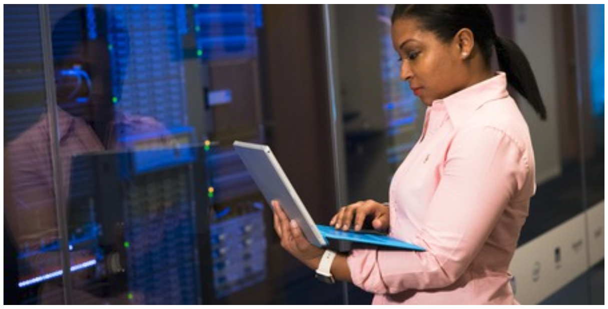
Career Profile



In movies and in the media, computer hackers are often portrayed as the bad guys—criminals who steal money or important information. What if you could be a good hacker? Somebody whose job is to find security flaws in computer systems; but rather intentionally try to break into… Read more

Security Incident Responder *In Demand!*

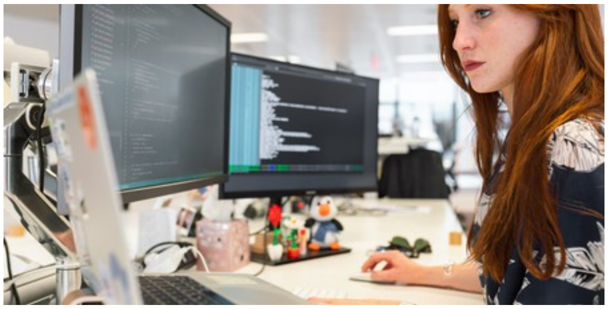
Career Profile



Security incident responders, also called intrusion analysts or incident response engineers, are like the "firefighters" of the cyber world. Companies can take steps to safeguard their computer networks and systems, but sometimes prevention is not enough and cyber attacks still happen. Sensitive data like customer credit card information can be stolen, entire websites could be brought down or altered, or personal contact information can be leaked. When this happens, incident responders must act… Read more

Cryptographer

Career Profile



Cryptographers, also called cryptologists and cryptanalysts, develop the encryption algorithms that keep our modern online transactions, like emails and credit card purchases, safe from prying eyes. Even if information or a message is stolen, as long as it is encrypted, the person who stole it cannot read it! Cryptographers also work to test and break these algorithms, to check them for weaknesses and vulnerabilities. They even analyze and decipher codes used by terrorists and foreign… Read more

Related Links

- Computer Science Project Ideas
- My Favorites

News Feed on This Topic

New Cybersecurity Projects for Students, *Science Buddies Blog*, October 3, 2017

Researchers develop a programme to find cipher vulnerabilities, *EurekAlert!*, April 23, 2021

New two-step algorithm could prove "a paradigm shift" in cloud data confidentiality, *EurekAlert!*, June 23, 2021

1 2 3 4 5 ... 34 >

Note: A computerized matching algorithm suggests the above articles. It's not as smart as you are, and it may occasionally give humorous, ridiculous, or even annoying results! Learn more about the News Feed

Cite This Page

General citation information is provided here. Be sure to check the formatting, including capitalization, for the method you are using and update your citation, as needed.

MLA Style

Finio, Ben. "Crack the Code: Breaking a Caesar Cipher." *Science Buddies*, 20 Nov. 2020, https://www.sciencebuddies.org/science-fair-projects/project-ideas/Cyber_p005/cybersecurity/crack-caesar-cipher. Accessed 4 Mar. 2022.

APA Style

Please Give Us Feedback!

Please rate the overall *quality* of this Project Idea

○ Excellent
○ Very Good
○ Good
○ OK
○ Poor

Do you think that you will do a project similar to this one?

○ Yes
○ Maybe
○ No

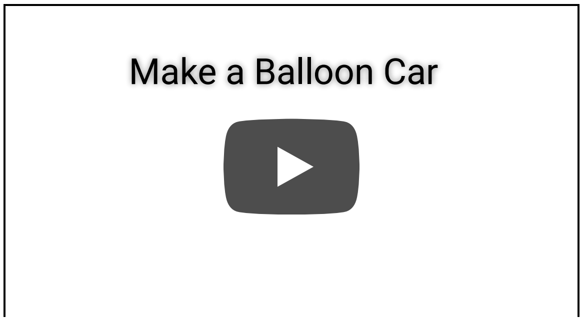
Send Feedback

Last edit date: 2020-11-20

Explore Our Science Videos

Harvest Water from Fog Science Project

How to Make an Archimedes Screw - STEM Activity



COMPANY

About Us
Sponsors
Partners
Contact Us
Work for Us
Image Credits
Site Map

PROJECT HELP

Science Fair Project Guide
Engineering Design Project Guide
Advanced Project Guide
Science Projects
Ask an Expert

GET INVOLVED

How to Donate
How to Volunteer

FIND US HERE

Join us on Facebook
Follow us on YouTube
Follow us on Twitter
Follow us on Pinterest
Email Us