Computer and Network Security: Self Study Lab

Lab #1 Basic Cipher with Unix

1. Introduction

The learning objective of this self-study lab is for students to get familiar with the concepts and practical usage in basic cipher techniques such as Caesar cipher and Substitution cipher.

> *The art of writing secret messages – intelligible to those who are in possession of the key and unintelligible to all others – has been studied for centuries. The usefulness of such messages, especially in time of war, is obvious; on the other hand, their solution may be a matter of great importance to those from whom the key is concealed. But the romance connected with the subject, the not uncommon desire to discover a secret, and the implied challenge to the ingenuity of all from who it is hidden have attracted to the subject the attention of many to whom its utility is a matter of indifference.*

Abraham Sinkov
In Mathematical Recreations & Essays
By W.W. Rouse Ball and H.S.M. Coxeter, c. 1938

2. Caesar cipher

Background: The Caesar cipher is named after the Roman military and political leader Gaius Julius Caesar (100 BC – 44 BC). Caesar used this relatively simple form of ciphering to encode military messages.

The classic version uses the capital letters `A-Z`, but, in principle, an arbitrary alphabet can be used. The first step is to write the alphabet down two times.

```
Input:   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
Output:  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
```

Now, the bottom alphabet is shifted by an arbitrary number of positions. The number of positions is the key-value. Shifting the bottom alphabet 3 positions to the right yields the following result:

```
Input:   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
Output:  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
```

The letter `A` becomes the letter `D`. `B` is replaced by `E` and `C` replaced by `F`, etc. The word "`EXAMPLE`" would be encoded by: "`HADPSOH`". You can play online Caesar cipher at https://www.cryptool.org/en/cto-ciphers/caesar

ROT13 ("rotate by 13 places", sometimes hyphenated ROT-13) is a simple letter substitution cipher that replaces a letter with the 13th letter after it, in the alphabet. ROT13 is a special case of the Caesar cipher, developed in ancient Rome.

Lab 1.1: Fill the ROT13 output to create a lookup table for the following input

```
Input:   ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz
Output:  NOPQRSTUVWXYZABCDEFGHIJKLMnopqrstuvwxyzabcdefghijklm
```

Security: The key length is identical to the size of the given alphabet. Using the capital letters, `A-Z` as alphabet allows 26 different keys, with the 26th key rendered meaningless because it would map each letter to itself. With only 25 meaningful keys, it would be quite easy to test for all possible keys until the correct one is found (brute-force analysis). The Caesar cipher can also easily be cracked with a frequency-analysis.

Linux tool for Caesar Cipher: The `tr` is an UNIX utility for translating, or deleting, or squeezing repeated characters. It will read from STDIN and write to STDOUT.

The syntax of `tr` command is:

```
$ tr [OPTION] SET1 [SET2]
```

Recommended Reading:

- http://www.linuxjournal.com/article/2563
- https://www.linuxnix.com/16-tr-command-examples-in-linux-unix/

Lab 1.2: The ROT13 is fairly easy to implement in Unix. Use the following command to encode the string "The Quick Brown Fox Jumps Over The Lazy Dog" in ROT13, and record the output.

```
echo "The Quick Brown Fox Jumps Over The Lazy Dog" | tr 'A-Za-z' 'N-ZA-Mn-za-m'
Ans : Gur Dhvpx Oebja Sbk Whzcf Bire Gur Ynml Qbt
```

Lab 1.3: What is the command to decode the output from Lab 1.2? Show the command and its output.

```
echo "Gur Dhvpx Oebja Sbk Whzcf Bire Gur Ynml Qbt" | tr ’N-ZA-Mn-za-m’ 'A-Za-z'
Ans: The Quick Brown Fox Jumps Over The Lazy Dog
```

Lab 1.4: Fill the ROT47 output to create a lookup table for the following input

```
Input:  !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNO
        PQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
Output:PQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~!"#$%&'()*+,-./
0123456789:;<=>?@ABCDEFGHIJKLMNO
```

Lab 1.5: What is the command to encode string "The Quick Brown Fox Jumps Over the Lazy Dog" in ROT47

```
echo "The Quick Brown Fox Jumps Over The Lazy Dog" | tr '\!-~' 'P-~\!-O'
```

3. Substitution cipher

Background: A simple substitution cipher is a method of concealment that replaces each letter of a plaintext message with another letter. Here is the key to a simple substitution cipher:

```
Input:    abcdefghijklmnopqrstuvwxyz
Output:   EKMFLGDQVZNTOWYHXUSPAIBRCJ
```

The key gives the correspondence between a plaintext letter and its replacement ciphertext letter. Using this key, every plaintext letters a would be replaced by ciphertext E, every plaintext letter e by L, etc. The plaintext message `simple substitution cipher` would become `SVOHTL`

SAKSPVPAPVYW MVHQLU.

Security: Solving the message not knowing the key is called **cryptanalysis**. If the cryptanalyst knew that the method of encryption was simple substitution cipher, then the cryptanalyst could try all possible keys to solve the message. Or, maybe not! How many keys are possible? How long would it take to try them all? When constructing a key for a simple substitution cipher, there are 26 choices of letters to substitute for a, then 25 remaining letters that can be substituted for b, then 24 remaining letters that can be substituted for c, etc. This results in

26x25x24x23x ... 3x2x1 = 26!

possible keys. That's a lot of keys; in fact, there are

403,291,461,126,605,635,584,000,000

keys.

Frequency Analysis: Frequency analysis is the study of letters or groups of letters contained in a ciphertext in an attempt to partially reveal the message. The English language (as well as most other languages) have certain letters and groups of letters appear in varying frequencies.

Knowing the usual frequencies of letters in English communication, if the encryption method does not effectively mask these frequencies it is possible to statistically determine parts of the plaintext from looking at the ciphertext alone. Let’s look at an example based on a plaintext encrypted with the Caesar Cipher – a cipher that provides no protection from frequency analysis.

Reading recommendation for frequency analysis:

- https://en.wikibooks.org/wiki/Cryptography/Frequency_analysis
- http://crypto.interactive-maths.com/frequency-analysis-breaking-the-code.html

Lab 2.1: Let’s the following are the simple substitution cipher with key:

Plaintext letters:	abcdefghijklmnopqrstuvwxyz
Ciphertext letters:	HUFRCOGMTZXLKPNWYVABQSIEDJ

Decrypt the message: BMC XTP MHBM PNBC NO HLL BMHB BMCD TPBCPR, UD TPBCVFCBTNP IMTFM BMCD RVCHK PNB NO.

THE KIN HATH NOTE OF ALL THAT THEY INTEND, BY INTERCETION WHICH THEY DREAM NOT OF

Lab 2.2: Suppose that the key for substitution cipher is unknown. Try to decrypt the message iq ifcc vqqr fb rdq vfllcq na rdq cfjwhwz hr bnnb hcc hwwhbsqvqbre hwq vhlq

Hint: Open **http://substitution.webmasters.sk/simple-substitution-cipher.php** for supporting tool

we will meet in the middle of the library at noon all arrangements are made

4. Base64

Background: Base64 is a group of similar binary-to-text encoding schemes that represent binary data in an ASCII string format by translating it into a radix-64 representation. The term Base64 originates from a specific MIME content transfer encoding. Each base64 digit represents exactly 6 bits of data. Three 8-bit bytes (i.e., a total of 24 bits) can therefore be represented by four 6-bit base64 encodings.

Essentially each 6 bits of the input is encoded in a 64-character alphabet. The "standard" alphabet uses A-Z, a-z, 0-9 and + and /, with = as a padding character.

The Base64 is not an encryption scheme since its encoding and decoding are public functions that anyone can evaluate. The only way that Base64 is related to cryptography is that it is convenient to encode ciphertext from some cryptosystem, which is uniformly distributed in 8-bit strings, in a limited set of US-ASCII that will not be munged or rejected in contexts that are limited to plain text, such as XML.

Reading recommendation for Base64:

- https://en.wikipedia.org/wiki/Base64
- https://www.lifewire.com/base64-encoding-overview-1166412

Linux tool for Base64: The `base64` is an UNIX utility for encoding and decoding data and print to standard output. The syntax of `base64` command is:

```
$ base64 [OPTION]… [FILE]
```

Lab 3.1 Playing the Base64 encoding and decoding at https://codebeautify.org/base64-encode

Lab 3.2 Encode the text input with Base64 using `base64`

```
$ echo -n 'Linux is rock!!' | base64
TGludXggaXMgcm9jayEh
```

Lab 3.3 Show the command and output for decoding the `base64` result from Lab 3.2

```
$ echo -n 'TGludXggaXMgcm9jayEh' | base64 -d
Linux is rock!!
```

Lab 3.4 Run the following command and compare the output.

```
$ echo -n '123456' | base64
$ echo -n '1234567' | base64
$ echo -n '12345678' | base64
$ echo -n '123456789' | base64
$ echo -n '1234567890' | base64
$ echo -n '12345678901' | base64

MTIzNDU2
MTIzNDU2Nw==
MTIzNDU2Nzg=
MTIzNDU2Nzg5
MTIzNDU2Nzg5MA==
MTIzNDU2Nzg5MDE=
```

Lab 3.5 What is the meaning of “=” and “==”

ขยายขนาดสตริงให้ได้ตาม **form**

Lab 3.6 Check the type of Linux command `mkdir` in `/bin` using file, then encode the `mkdir` file using `base64` and see the output.

```
$ file /bin/mkdir
$ base64 /bin/mkdir > mkdir-base64.txt
```

Lab 3.7 Show the command to decode the image `pix.base64` from https://goo.gl/EkPjkr
Who is he?

```
base64 --decode pix.base64 > pix.png
อ. อนันต์
```

5. Recommended Reading:

1. **Simple Substitution Ciphers:** https://www.nku.edu/~christensen/1402%20Simple%20substitution%20and%20Caesar%20cipher.pdf
2. **Understanding Linux / UNIX tr command:** https://www.cyberciti.biz/faq/how-to-use-linux-unix-tr-command/
3. **Frequency Analysis Tools:** http://www.counton.org/explorer/codebreaking/frequency-analysis.php
4. **Classical Ciphers and Frequency Analysis Examples** https://sandilands.info/sgordon/classical-ciphers-frequency-analysis-examples