Introduction to Cryptography

Target Course

CS1

Learning Goals

A student shall be able to:

1. Be able to describe the uses of cryptography to support cybersecurity goals and foundational concepts

IAS Outcomes

The CS2013 Information Assurance and Security outcomes addressed by this module are:

IAS Knowledge Topic	Outcome
Cryptography	1. Describe the purpose of Cryptography and list ways it is used in data communications. 2. Define the following terms: Cipher, Cryptanalysis, Cryptographic Algorithm, and Cryptology and describe the two basic methods (ciphers) for transforming plain text in cipher text.

Dependencies

- Familiarity with modulo arithmetic
- Assessments require knowledge of loops, conditionals and either strings or arrays

Summary

Use exceptions to validate input data.

Estimated Time

[Provide the estimated amount of lecture time to cover this module, using the notion of time as defined in CS2013.]

Summary

Introduce data encryption (cryptography) by simple encryption schemes including substitution and transposition schemes.

Materials

How can two parties communicate confidentially over an insecure channel?
Consider the simple scenario where Alice and Bob want to communicate confidentially over an insecure channel. Alice wants to send Bob a message, but she cannot simply transmit it in **plaintext** since eavesdroppers (like Eve) on the channel can then read the message. Instead, Alice uses an encryption algorithm to convert plaintext to an encrypted form called ***ciphertext*** and transmits this to Bob. Once Bob receives the ciphertext he uses a decryption algorithm to recover the original plaintext message.

The encryption and decryption algorithms are based on some secret information that is only known to Alice and Bob. For the communication to be confidential it must be infeasible for a person without the knowledge of these secrets to determine the plaintext from the ciphertext. The encryption and decryption algorithms make up the **Cryptographic algorithms**.

More formally, let p be the plain text message, E and D be the encryption and decryption algorithms, and c be the ciphertext. Let k_1 and k_2 be the secret keys that Alice and Bob use. The communication between Alice and Bob looks like this:

Alice	Insecure Channel	Bob
1. Encrypt plaintext p, $c = E(p, k_1)$ 2. Transmit c over channel	c →	3. Decrypt c to recover the plaintext $p = D(c, k_2)$

The cryptographic algorithms (E, D) take the secret key as an input. This means that the algorithms themselves can be publically available software.

The cryptographic algorithms and the secret keys must somehow line up together so that Bob is eventually able to recover the plain text message. That is, it must be true that $p = D(E(p,k_1), k_2))$. In certain cryptographic algorithms the same secret key is used for encryption and decryption (i.e. $k_1 = k_2$). In other algorithms the keys are different, k_1 is published and k_2 is kept secret and only known to Bob.

Cryptography is the study of designing cryptographic algorithms to keep messages secure. **Cryptanalysis** is the study of methods for cracking cryptographic algorithms without knowledge of secret keys. The practitioners of cryptography and cryptanalysis are cryptologists. Modern cryptologists are generally trained in theoretical mathematics and theoretical computer science.

What are some basic substitution based encryption schemes?

Caesar's cipher or **Caesar shift**, is one of the simplest encryption schemes. It is a *substitution cipher* in which each letter in the plaintext is substituted by a letter some fixed number of positions down the alphabet, wrapping around. For example, with shift of 3, `A` in the plaintext would be replaced by `D`, `B` by `E`, and so on wrapping around so that `X` would be replaced by `A` etc., as shown below:

```
Plain:    ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher:   DEFGHIJKLMNOPQRSTUVWXYZABC
```

The shift value (3 above) is the ***key*** of the encryption scheme. Encryption is done by shifting each letter of the plain text right by this key and decryption is done by shifting each letter of the ciphertext left by the same key.

To implement the Caesar cipher first let m be the size of the alphabet. For example, if the messages being transmitted contain only capital letters A,…Z, then $m=26$. Now transform all letters in the alphabet into numbers. For example, `A = 0, B = 1,..., Z = 25`. Encryption of a plaintext letter x by a shift k is described using modulo arithmetic as

$$E(x, k) = (x+k) \bmod m$$

Using modulo arithmetic ensures that the encryption wraps around the alphabet after the shift so that the result will be in the range 0, …25. Thus the cipher text will only contain the letters from the original alphabet. Decryption is described similarly as

$$D(x, k) = (x- k) \bmod m$$

(Another way to understand the modulo operations above is if the shift results in a value outside the range [0,25] then 26 must be added or subtracted from the result. That is, if $(x+k) > 25$, subtract 26; if $(x-n) < 25$, add 26.)

Example: Use capital letters, so $m=26$ and the plain text to encrypt is `ATTACKTODAY`

Plaintext	A	T	T	A	C	K	T	O	D	A	Y
plain x	0	19	19	0	2	10	19	14	3	0	24
Encryption:	3	22	22	3	5	13	22	17	6	3	1

x + 3 mod 26 (= cipher x)											
Ciphertext	D	W	W	D	F	N	W	R	G	D	B
Decryption x - 3 mod 26 (= plain x)	0	19	19	0	2	10	19	14	3	0	24

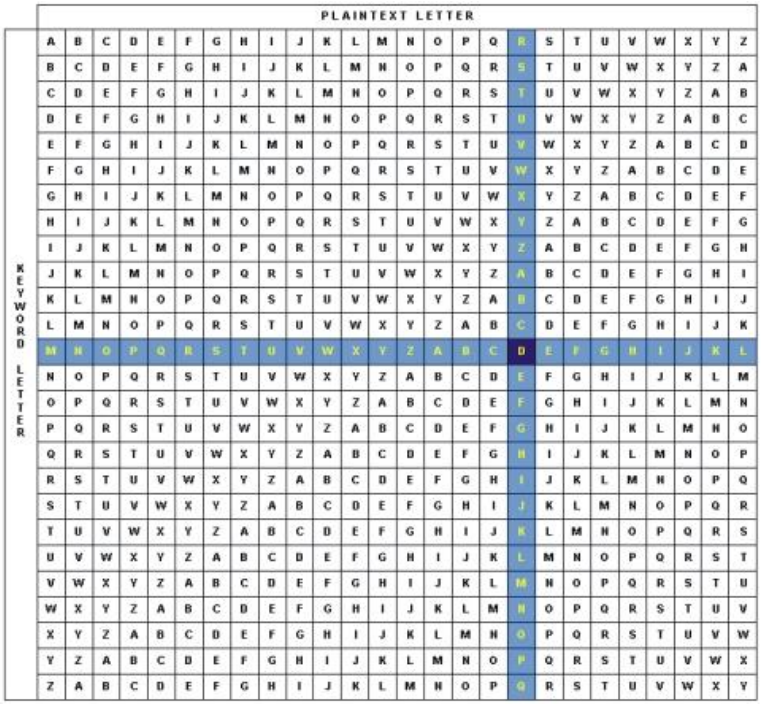
The **Affine Cipher** is a related encryption scheme which also maps letters in an alphabet to a numeric, and encrypts the plaintext using a simple mathematical function. In an affine cipher the encryption function for a single letter x is $E(x, a, b) = (ax + b) \mod m$.

The decryption function is $D(x, a, b) = a^{-1} (x - b) \mod m$, where a^{-1} is the modular multiplicative inverse of a. The secret keys in this scheme are a and b. An additional requirement is that a must be co-prime with m for the modular multiplicative inverse of a to exist. *Note that using a=1 results in the Caesar Cipher.*

The **Vigenère cipher** uses a series of different Caesar ciphers based on the letters of a keyword. The cipher uses a table with 26 rows where each row corresponds to one of 26 possible Caesar ciphers. The encryption uses a *keyword.* At different points in the encryption process, the cipher looks up the row (corresponding to a letter from repeating the keyword), and the column (corresponding to a letter being encrypted). The plaintext letter is replaced by the character in the table at position `table[row, col]`.

Example: Note that the key LEMON is repeated across the entire plaintext message

Plaintext: `ATTACKATDAWN`
Key: `LEMONLEMONLE`
Ciphertext: `LXFOPVEFRNHR`



What are transposition based encryption schemes?

Transposition ciphers are based on shuffling around the order of characters. In a simple columnar transposition cipher, for encryption the plaintext is written onto a table with rows of fixed sized (see example below) and read out column by column, where columns are chosen in some scrambled order. A keyword can be used to define the size of the rows and the scrambling order for columns. For example, suppose we are encrypting `WE ARE DISCOVERED`.

FLEE AT ONCE, and the key word is ZEBRAS. Since ZEBRAS is of length 6 the size of each row is 6. The table is filled in completely by adding random letters in the last row of the table (see the letters Q K J E and U in the last row below). The scrambling of columns is defined by the alphabetical order of the letters in the keyword. In this case, the order is "6 3 2 4 1 5". Reading off the column by column in the scrambled order gives us the ciphertext.

```
Plaintext:  WE ARE DISCOVERED. FLEE AT ONCE
Keyword: ZEBRAS
Row length = 6

Encryption table:
6 3 2 4 1 5
W E A R E D
I S C O V E
R E D F L E
E A T O N C
E Q K J E U
Column scrambling: 632415

Ciphertext: EVLNE ACDTK ESEAQ ROFOJ DEECU WIREE
```

To decrypt the cipher text the receiver has to calculate column lengths by dividing the message length by the key length (e.g., column length is 5 = 30/6). Then the ciphertext is written out in columns of this length (see the decryption table below). Finally the columns are un-scrambled by reordering the columns according the ordering required to get the alphabetized keyword letter back to its correct position in the keyword. For example, the ordering required to get ABERSZ (i.e., the keyword with its letters in alphabetical order) back to the word ZEBRAS is "5 3 2 4 6 1". After unscrambling columns the plaintext can be read off by reading row by row. In the example, reading row 1 using the column reordering gets us W E A R E D, which are the first six letters of the plaintext.

```
Ciphertext: EVLNE ACDTK ESEAQ ROFOJ DEECU WIREE
Keyword: ZEBRAS
Column length = 5

Decryption table:
5 3 2 4 6 1
E A E R D W
V C S O E I
L D E F E R
N T A O C E
E K Q J U E
Column scrambling: 5 3 2 4 6 1
```

There are various ways to make transposition ciphers more complicated, e.g. putting the text through another transposition, and making the read out rule more complicated e.g. read by "spiral inwards, clockwise, starting from the top right", etc.

How secure are the encryption schemes described above?

The Caesar Cipher is easy to crack using straightforward application of frequency analysis. For instance, if P is the most frequent letter in a ciphertext one might suspect that P corresponds to E, because E is the most frequently used letter in English. Additionally one can try to decrypt the cipher text using all possible shifts values in the range of the size of the alphabet e.g. from $0, \ldots m$. If this range size is small (as 26 in our example above) it is easy and quick to try all

possible decryptions using a computer program. Similar cryptanalysis techniques can be used to break the Affine cipher.

The Vigenère cipher is harder to crack. For three centuries it resisted all attempts to break it earning it the title **le chiffre indéchiffrable** (French for 'the indecipherable cipher'). Today there are well known methods of breaking the Vigenère cipher. Simple frequency analysis does not work: E can be encrypted as different ciphertext letters at different points in the message. The primary weakness of the Vigenère cipher is the repeating nature of its key. Some cryptanalysis methods are based on correctly guessing the key's length, then treating the cipher text as interwoven Caesar ciphers, which are individually broken.

The main vulnerability of transposition ciphers is that the letters of the ciphertext are the same as those of the plaintext, a frequency analysis on the ciphertext would reveal that each letter has approximately the same likelihood as in English. Using these clues cryptanalysts can then use a variety of techniques to determine the right ordering of the letters to obtain the plaintext. Computers can break almost all transposition ciphers even those using more complicated rules.

How does modern encryption schemes work (since the ones described above are so easy to break)?

Modern cryptography is heavily based on theoretical mathematics and computer science. They rely on problems from number theory which are believed to be hard to solve computationally e.g. Integer factorization, discrete logarithms and elliptical curves. Cryptographic algorithms are designed around computational hardness assumptions which essentially claim that while it is theoretically possible for an adversary to crack the encryption schemes, it is computationally impractical to do so (for example it would take more than 1000 years to crack one cipher). Modern schemes are termed **computationally secure** so improvements in algorithms for the underlying hard problems (e.g. integer factorization), and faster computing technology require continually adapting these solutions.

What security goals and concepts does Cryptography support?

Traditionally, the purpose of cryptography was to allow two parties (Alice and Bob) to communicate confidentially over an insecure channel which is subject to eavesdropping by adversaries (Eve). Although the encrypted messages between Alice and Bob are visible to Eve, as the messages are encrypted Eve is be unable to make sense of them. Today cryptography continues to be used for secure communication and also to store sensitive data such as passwords securely. In these contexts cryptography supports the security goal of confidentiality, providing defense mechanisms against eavesdroppers.

Additionally it is also used to digitally prove identity (digital signatures), time stamp electronic documents to prove delivery by at certain time, electronic money and to protect copyrighted materials. In these context it is used to uphold authenticity, providing detection and deterrence of attacks.

Finally, cryptography should be used to persistently store any data that is deemed sensitive or private, including health and financial records.

Assessment Methods

Possible programming assignments

- Write a program to encrypt and decrypt messages using Caesar Cipher.
- Write a program that takes ciphertext encrypted with Caesar Cipher and decrypts it using all possible shift values. Each decryption string is output to the screen.
- Write a program to encrypt and decrypt messages using the column transposition scheme.

References:

[1] **https://en.wikipedia.org/wiki/Caesar_cipher**

[2] https://en.wikipedia.org/wiki/Affine_cipher

[3] https://en.wikipedia.org/wiki/Transposition_cipher

[4] https://en.wikipedia.org/wiki/Cryptography#Symmetric-key_cryptography

[5] Bruce Schneier – Applied Cryptography

[6] Goodrich and Tomassia – Introduction to computer security