CS 6324: Information Security **Spring 2021**

Homework #1

Due date & time: 11:59pm CST on February 12, 2021. Submit to eLearning by the due time.

Additional Instructions: The submitted homework must be typed. Using LATEX is recommended, but not required.

Problem 1 (10 pts) Confidentiality, Integrity, Availability.

- (3 pts) State what is Confidentiality, Integrity, and Availability.
- (3 pts) For each, give two examples where they are violated.
- (4 pts) Identify two computer security control measures on your computer(s). Which of the three properties Confidentiality, Integrity, and Availability do they aim at providing? What kinds of adversaries they **cannot** defend against?

Answer.

- (3 pts) State what is Confidentiality, Integrity, and Availability.
  1. **Confidentiality** is the property that information should only be disclosed to authorized users.
  2. **Integrity** is the property that information should not be modified without that modification being known.
  3. **Availability** is the property that the information should be accessible when needed.
- (3 pts) For each, give two examples where they are violated.
  1. **Confidentiality**
     • spyware on a system allowing an outside party to monitor/access confidential data.
     • an sql injection attack on a database which dumps the whole database
     • man in the middle attack which monitors network traffic
  2. **Integrity**
     • malware on a system that modifies certain data on your system, for instance it could add malware to other executables without you knowing
     • man in the middle attack which modifies network traffic without either party knowing.
     • the stuxnet attack which reported false readings from the scada control units, causing the uranium enrichment devices to malfunction.

3. **Availability**
   • malware which deletes the files off your system. Or ransomware which encrypts your hard drive, rendering your system unavailable.
   • denial of service attack against a website to prevent legitimate access to it
   • changing someones password so that they are unable to access a certain service.

- (4 pts) Identify two computer security control measures on your computer(s). Which of the three properties Confidentiality, Integrity, and Availability do they aim at providing? What kinds of adversaries they **cannot** defend against?

  1. **Hard drive encryption**
     • Confidentiality, Integrity
     • It makes the data confidential since it is not easy to decrypt the contents of the disk. And it leads to integrity because an attacker can not modify the contents of the disk without knowing the key to decrypt/encrypt.
     • This can not defend against adversaries who are already on your system. For instance, if you have malware on your system, then that malware will probably (at some point, if not always) be able to access the data on the encrypted hard drive. Hard drive encryption is only useful until you enter your key and decrypt it, then the OS allows access to the encrypted data and reads/writes can occur.
  2. **Anti Virus Software**
     • Confidentiality, Integrity, Availability
     • Because malware can undermine all of these security properties, and AV software is aimed to prevent malware, AV software can claim that it tries to uphold all three properties.
     • However, AV software doesn't work very well because it is based known signatures for the malware. So any polymorphic or encrypted code can sneak by AV detection, as does any malware that has never been seen before and only targets a small subset of people (ex. spear phishing 3 employees at a law firm with unknown malware). So while AV software can work well against naive large scale malware attacks, more targeted and sophisticated attacks are more difficult to handle.

∎

Problem 2 (30 pts) Probability Review.

1. (10 pts) We roll two fair 6-sided dice.
   (a) Find the probability that doubles are rolled.
   (b) Given that the roll results in a sum of 6 or less, find the conditional probability that doubles are rolled.
   (c) Find the probability that that larger of the two die's outcomes is at least 4.
   (d) Given that the two dice land on different numbers, find the conditional probability that at least one die roll is a 1.
   (e) Let X be the event that the first die results in an odd number, and Y be the event that the rolled sum is an even number. Compute $\mathbf{Pr}[X], \mathbf{Pr}[Y], \mathbf{Pr}[X \wedge Y]$. Are X and Y independent?

(f) Let X be the event that the first die results in a multiple of 2, and Y be the event that the rolled sum is a multiple of 2. Compute $\mathbf{Pr}[X], \mathbf{Pr}[Y], \mathbf{Pr}[X \wedge Y]$. Are X are Y independent?

(g) Let X be the event that the first die results in a multiple of 3, and Y be the event that the rolled sum is a multiple of 2. Compute $\mathbf{Pr}[X|Y]$ and $\mathbf{Pr}[Y|X]$. Are X are Y independent?

Answer.

(a) (1 pts) $\frac{1}{6}$

(b) (1 pts) $\frac{1}{5}$

(c) (1 pts) $\frac{6\times 6-3\times 3}{6\times 6} = \frac{3}{4}$

(d) (1 pts) $\frac{1}{3}$

(e) (0.5 pts) $\mathbf{Pr}[X] = 1/2$
(0.5 pts) $\mathbf{Pr}[Y] = 1/2$
(0.5 pts) $\mathbf{Pr}[X \wedge Y] = 1/4 = \mathbf{Pr}[X]\mathbf{Pr}[Y]$.
(0.5 pts) X and Y are independent.

(f) (0.5 pts) $\mathbf{Pr}[X] = 1/2$
(0.5 pts) $\mathbf{Pr}[Y] = 1/2$
(0.5 pts) $\mathbf{Pr}[X \wedge Y] = 1/4 = \mathbf{Pr}[X]\mathbf{Pr}[Y]$.
(0.5 pts) X and Y are independent.

(g) (0.5 pts) $\mathbf{Pr}[X|Y] = 1/3$
(0.5 pts) $\mathbf{Pr}[Y|X] = 1/2$
(1 pts) $\mathbf{Pr}[X] = 1/3$. $\mathbf{Pr}[Y] = 1/2$. $\mathbf{Pr}[X \wedge Y] = 1/6 = \mathbf{Pr}[X]\mathbf{Pr}[Y]$. X and Y are independent.

■

2. (5 pts) Let X denote the event that a student pass the midterm exam in a course. And Y denote the event that the student pass the final exam. We know that $\mathbf{Pr}[X] = 0.75$, $\mathbf{Pr}[Y] = 0.8$, and $\mathbf{Pr}[Y|X] = 0.9$. (The probability that a student will pass the final exam given that he or she has passed the midterm exam is 0.9.) Compute $\mathbf{Pr}[X \wedge Y]$, $\mathbf{Pr}[\neg X|Y]$, and $\mathbf{Pr}[Y|\neg X]$. Are X and Y independent?

Answer. (1 pts) $\mathbf{Pr}[X \wedge Y] = \mathbf{Pr}[X]\mathbf{Pr}[Y|X] = 0.675$
(1 pts) $\mathbf{Pr}[\neg X|Y] = 1 - \mathbf{Pr}[X|Y] = 1 - \mathbf{Pr}[X \wedge Y]/\mathbf{Pr}[Y] = 0.15625$
(1 pts) $\mathbf{Pr}[Y|\neg X] = \mathbf{Pr}[\neg X|Y]\mathbf{Pr}[Y]/\mathbf{Pr}[\neg X] = 0.5$
(2 pts) $\mathbf{Pr}[X \wedge Y] \neq \mathbf{Pr}[X]\mathbf{Pr}[Y]$. X and Y are not independent.

■

3. (6 pts) There are 78 qualified applicants for teaching positions in an elementary school, of which some have at least five years' teaching experience and some have not, some are married and some are single, with the exact breakdown being

	Married	Single
At least five years teaching experience	18	12
Less than five years teaching experience	30	18

(a) The order in which the applicants are interviewed is random. M is the event that the first applicant interviewed is married and F is the event that the first applicant interviewed has at least five years teaching experience. Find the following probabilities: $\mathbf{Pr}[M], \mathbf{Pr}[F], \mathbf{Pr}[M \wedge F], \mathbf{Pr}[M|F], \mathbf{Pr}[F|M]$. Are M and F independent?

(b) Suppose that there is only one opening in the third grade, and each applicant with at least five year experience has *twice* the chance of an applicant with less than five year experience. Let U denote the event that the job goes to one of the single applicants, and V the event that it goes to an applicant with less than five year experience. Find the following probabilities: $\mathbf{Pr}[U], \mathbf{Pr}[V], \mathbf{Pr}[U \wedge V], \mathbf{Pr}[U|V], \mathbf{Pr}[V|U]$. Are U and V independent?

Answer.

(a) (0.5 pts each)
$\mathbf{Pr}[M] = 8/13$
$\mathbf{Pr}[F] = 5/13$
$\mathbf{Pr}[M \wedge F] = 3/13$
$\mathbf{Pr}[M|F] = 3/5$
$\mathbf{Pr}[F|M] = 3/8$
Are M and F independent? No

(b) (0.5 pts each)
$\mathbf{Pr}[U] = 7/18$
$\mathbf{Pr}[V] = 4/9$
$\mathbf{Pr}[U \wedge V] = 1/6$
$\mathbf{Pr}[U|V] = 3/8$
$\mathbf{Pr}[V|U] = 3/7$
Are U and V independent? No

■

4. (9 pts) A test for a certain rare disease is assumed to be correct 95% of the time. Let S denote the event that a person has the disease, and T denote the event that the test result is positive. We have $\mathbf{Pr}[T|S] = 0.95$ and $\mathbf{Pr}[\neg T|\neg S] = 0.95$. Assume that $\mathbf{Pr}[S] = 0.001$, i.e., the probability that a randomly drawn person has the disease is 0.001.

   (a) Compute $\mathbf{Pr}[S|T]$, the probability that one has the disease if one is tested positive.

   (b) Assume that one improves the test so that $\mathbf{Pr}[T|S] = 0.998$, i.e., if one has the disease, the test will come out positive with probability 0.998. Other things remain unchanged. Compute $\mathbf{Pr}[S|T]$.

   (c) Assume that another improvement on the test improves $\mathbf{Pr}[\neg T|\neg S] = 0.998$, but have $\mathbf{Pr}[T|S]$ remain at 0.95. Compute $\mathbf{Pr}[S|T]$.

Answer.

1. (3 pts)0.01866
2. (3 pts)0.01958
3. (3 pts)0.32225

Problem 3 (8 pts) Cryptanalysis Concepts.

- (3 pts) Explain what do ciphertext-only attacks, known-plaintext attack, and chosen plaintext attack mean?

  Answer. Ciphertext-only attack is when an attacker only knows cipher texts, no plaintext. Known-plaintext attack is when an attacker has cipher/plain text pairs that are given to them. Chosen plaintext attack is when the attacker can choose the plain text, and get the corresponding cipher text.

- (2 pts) What attack can be used to break the substitution cipher under a ciphertext-only attack? Explain the simplest way to break it under a known-plaintext attack?

  Answer. Frequency analysis. To break it under a known-plaintext attack you can just do a lookup on the plaintext-¿encrypted text mappings and get the offsets directly.

- (3 pts) Explain how one may be able to carry out a known-plaintext attack against the wireless encryption between the laptop used by the target user and a wireless access point. Explain how one may be able to carry out a chosen-plaintext attack.

  Answer. Known plain text can occur if you are able guess the IP and protocol of someone on the network, then you can use that as a partial known plain text for the message. Even better, if you can monitor traffic external to the wifi, then you can use timestamps to match up full messages to the packets you sniff off the air.

  To perform a chosen plaintext attack, if you know the IP address for a machine, you could send that machine messages. Like pinging it, or any other message you choose to send.

Problem 4 (8 pts) Consider the following "Double Vigenère encryption". We choose two random keys K_1 and K_2 of lengths ℓ and $\ell + 1$. To encrypt a message, we first use key K_1 to encrypt in the Vigenère fashion, then use K_2 to encrypt.

- (4 pts) Describe how to break this encryption scheme under a ciphertext only attack. Does this double encryption offer increased level of security over Vigenère encryption against a ciphertext only attack?

  Answer. The effect of this cipher is such that the plaintext is encrypted under $K_1^+ \oplus K_2^+$, where K^+ denotes repeating the key K for as many times as needed. The stream $K_1^+ \oplus K_2^+$ repeats every $\ell(\ell+1)$. Thus one can attack it as a Vigenère cipher with key length $\ell(\ell+1)$. This should offer increased level of security when compared with Vigenère encryption with key length ℓ, because frequency analysis is made more difficult. If this is the most effective attack, then this also offers improved security over Vigenère with key length $2\ell + 1$ in ciphertext only attack. However, we cannot rule out a clever attack that causes the effective level of security equivalent to that of Vigenère with key length $2\ell + 1$.

- (4 pts) Suppose that we know that $10 \leq \ell \leq 19$. Given a ciphertext of length 50 and its corresponding plaintext, describe how to recover the keys so that any other message encrypted under

the same key pair can be decrypted.

Answer. For each possible key length ℓ, we have $2\ell + 1$ unknown letters for K_1 and K_2. We observe that given any pair (K_1, K_2), the key pair $(K_1 + x, K_2 - x)$ (each letter is added or subtracted by x) will have exactly the same effect. Therefore, there are only 2ℓ unknowns. Each letter in the plaintext/ciphertext pair gives us one linear equation: e.g., the first equation is $(m_1 + k_1 + k'_1) \mod 26 = c_1$. We have 50 equations, which is significantly more equations than the number of unknowns. For each ℓ from 10 to 19, construct the 50 linear equations for 2ℓ unknowns, then use the standard techniques for solving systems of linear equations to solve them. When ℓ is the actual key length, then there exists a unique solution. Here is a small example illustrating this, I use $\ell = 2$ and use arithmetic notation rather than modular arithmetic notation, if suffices to use the first four equations: $k_1+k'_1 = b$, $k_2+k'_2 = c$, $k_1+k'_3 = d$, $k_2+k'_1 = e$, set $k_1 = a$ for any a, we have $k_2 = a+e-b$, $k'_1 = b-a$, $k'_2 = c+b-a-e$, $k'_3 = d-a$. When we try with an incorrect ℓ, we are fitting 2ℓ variables to 50 linear equations and are over-constrained. The probability that a solution exists by chance is very small. (To establish this precisely is actually not easy, though the intuition should be clear.)

Problem 5 (14 pts) Consider the following enhancement of the Vigenère cipher. We assume that the plaintext is a case-insensitive English text using only the 26 letters (without space or any other symbol). To encrypt a plaintext of length n, one first uniformly randomly generates a string over the alphabet $[A..Z]$ of length 17, and then inserts this string into the beginning of the plaintext. That is, we first construct a string $x = x_1x_2 \dots x_{n+17}$, such that $x_1 \cdots x_{17}$ is the string we have generated, and $x_{18} \cdots x_{n+17}$ is the original plaintext string. We then construct a string y as follows: $y_i = x_i$ for $1 \le i \le 17$, and for $i \in [18, n + 17]$, y_i is the result of using y_{i-17} to encrypt x_i; that is, when the x_i's and y_i's are treated as numbers in $[0..25]$, we have $y_i \leftarrow ((x_i + y_{i-17}) \mod 26)$. We then apply the Vigenère cipher to the string y, while making sure that the key length is not a multiple of 17.

- Implement the encryption algorithm and decryption algorithm for this cipher in a programming language of your choice. Include the core part of your code.
  (4 pts) for encryption and (4 pts) for decryption.
- Choose a key, a plaintext, and run the encryption code multiple times, and ensure that the de-cryption results in the original plaintext. Include the key, the plaintext, and 3 ciphertexts.
- Write the Pseudo-code to recover enough information from a known (plaintext,ciphertext) pair to decrypt other messages encrypted under the same key. That is, the pseudo-code takes three inputs (M_1, C_1, C_2), where C_1 is a ciphertext of M_1, and outputs M_2, the message encrypted in C_2.
  (3 pts) **Answer Idea. Let $z_1 \cdots z_{n+17}$ denote the ciphertext, and $k_1 \cdots k_{n+17}$ denote the repeated key sequence. We observe the following relationship: for all $i \in [1, n]$, we have**

$$(y_i + k_i) \bmod 26 = z_i \wedge (x_{i+17} + y_i + k_{i+17}) \bmod 26 = z_{i+17}$$

We thus have

$$(z_{i+17} - z_i) \bmod 26 = (x_{i+17} + k_{i+17} - k_i) \bmod 26$$

Thus knowing both the z_i sequence and x_i sequence, we can compute the sequence $k'_{\mathrm{i}} = (k_{i+17} - k_i) \bmod 26$, which enables us to decrypt any new ciphertext sequence $w_1 \cdots w_{n+17}$ as $m_i = (w_{i+17} - w_i - k'_{\mathrm{i}}) \bmod 26$.

- How to effectively attack this cipher in a ciphertext only attack?
  (3 pts) **Answer Idea. Given a ciphertext sequence z_i, compute the sequence $c_i = z_{i+17} - z_i$, then c_i can be viewed as ciphertext from encryption with $k_{i+17} - k_i \bmod 26$, which repeats every ℓ. Thus one can attack this as ciphertext of Vigenère cipher using key length ℓ.**

Problem 6 (10 pts) Consider an example of encrypting the result of a 6-side dice (i.e., $M \in [1..6]$), as follows. Uniformly randomly chooses $K \in [1..6]$, ciphertext is $C = (M * K) \bmod 13$. The ciphertext space is thus $[1..12]$. We have $\mathbf{Pr}[\mathsf{PT} = 1] = \mathbf{Pr}[\mathsf{PT} = 2] = \mathbf{Pr}[\mathsf{PT} = 3] = \cdots = \mathbf{Pr}[\mathsf{PT} = 6] = 1/6$, and we use a vector notation $\mathbf{Pr}[\mathsf{PT}] = \langle 1/6, 1/6, \ldots, 1/6 \rangle$ to denote this.

- Assume that you stole a glance at the dice value and saw that there are many dots on it, and hence are quite certain that M is either a 5 or a 6. You then learned the ciphertext of the encrypted dice value. Under which ciphtertext value(s) can you learn the value M.
  Hint: You may want to start by writing out the 6 by 6 table of the ciphertext for each possible combination of plaintext and key.
  (3 pts) **Answer. When cipher text is : 2, 6, 7, 11, and you know plaintext is either 5 or 6, you can tell exactly which plaintext was input.**
- Compute $\mathbf{Pr}[\mathsf{PT}|\mathsf{CT} = i]$, for each $i \in [1..12]$, similar to the one for $\mathbf{Pr}[\mathsf{PT}]$ given above.
  (3 pts) **Answer.**
  The plain text is along the top. The first column, 1...12 is the cipher text value i.

	1	2	3	4	5	6
1	1	0	0	0	0	0
2	1/4	1/4	1/4	0	1/4	0
3	1/3	0	1/3	1/3	0	0
4	1/5	1/5	0	1/5	1/5	1/5
5	1/4	0	1/4	0	1/4	1/4
6	1/4	1/4	1/4	0	0	1/4
7	0	0	0	1/2	1/2	0
8	0	1/2	0	1/2	0	0
9	0	0	1	0	0	0
10	0	1/3	0	0	1/3	1/3
11	0	0	0	1/2	0	1/2
12	0	1/5	1/5	1/5	1/5	1/5

- Show that if K is uniformly randomly chosen from $[1..12]$, then this cipher provides perfect secrecy.

  (4 pts) **Answer.**
  Note that the key space has changed form $[1...6]$ to $[1...12]$. Working out the probabilities $\mathbf{Pr}[\mathsf{PT}|\mathsf{CT} = i]$ for all i = $[1...12]$, you will find that the vector is $< 1/6, 1/6, 1/6, 1/6, 1/6, 1/6 >$. This is the definition for perfect secrecy

Problem 7 (5 pts) Consider the following way of using the Vigenere cipher to send one encrypted message. The possible plaintexts are English texts of length 100. The key is a random string of length 50. Show that this does not satisfy perfect secrecy by finding two plaintext messages M_1, M_2 and a ciphertext message C such that:

$$\mathbf{Pr}[\mathsf{CT} = C \mid \mathsf{PT} = M_1] \neq \mathbf{Pr}[\mathsf{CT} = C \mid \mathsf{PT} = M_2]$$

Answer: Choose M_1 and C to be the repetition of a length-50 string, and M_2 to be a non-repeating string, then $\mathbf{Pr}[\mathsf{CT} = C \mid \mathsf{PT} = M_2] = 0$.

Problem 8 (5 pts) Prove that for any cipher that offers perfect secrecy, the number of the possible keys must be at least as large as the number of plaintexts.

Proof. If a cipher provides perfect secrecy, then given a ciphertext C_0, for any plaintext M, there must exist at least one key that encrypts M into C_0. (Otherwise, observing C_0 as the ciphertext leads one to conclude that M cannot be the plaintext.) Furthermore, given two different messages, the keys that encrypt them to C_0 must be different, as otherwise, one cannot decrypt the ciphertext C_0 into a unique plaintext message. Hence for every plaintext message, there must exist a different key. Hence the number of keys must be at least as large as the number of plaintext messages. ∎

Problem 9 (10 pts) Implement the following encryption/decryption function, which uses the RC4 stream cipher.

- byte[] encrypt(byte[] pt, byte[] key)
- byte[] decrypt(byte[] ct, byte[] key)

You should implement RC4 algorithm yourself. Google to find out the algorithm. You can assume that the key is an array of length between 16 and 32 bytes. You need to use a 256-bit (32-byte) initial vector so that when one invokes the encrypt function with the same pair of plaintext and key twice, with high probability the resulting ciphertexts are different.

You can choose any programming language to do this problem. You need to figure out what library function to call to generate a random IV. Note that the random IV is required to be unpredictable. This call to generate IV should be the only library call. You should drop the first 3072 bytes of RC4's output as recommended. You should run your code to verify that decrypt(encrypt(pt, key), key) = pt.

Include your code in the HW submission, and provide information regarding the library function you use to generate the random IV.

Answer. (2 pts) KSA

(2 pts) PRGA

(2 pts) Encryption / Decryption

(4 pts) IV. IV should be independent of key and be part of ciphertext (should not be an input for encryption/decryption function) ∎