Peyrin & Ryan
Summer 2020

CS 161
Computer Security

HW3

This practice exam was generated for `foo@bar.com`.

To prepare you for the midterm format, Homework 3 is formatted exactly like the midterm. Please write your answers on the practice answer sheet (or a blank piece of paper if you don't have printer access), and submit scans or pictures of the filled-out answer sheet to the "Homework 3" assignment on Gradescope.

Practice answer sheet link: https://cs161.org/assets/mt/practice_mt_answer_sheet.pdf

For questions with **circular bubbles**, you may select exactly *one* choice on the answer sheet.

○ Unselected option

● Only one selected option

For questions with **square checkboxes**, you may select *zero* or more choices on the answer sheet.

■ You can select

■ multiple squares

For questions with a **large box**, you need to provide justification in the blank space below the question on the answer sheet.

You have 1 week (110 minutes for the actual exam). There are 3 questions of varying credit (61 points total).

[NOTE: This homework has no instant feedback, so instead we will grade this homework out of 40 points. Anything above 40 points is full credit on the homework.]

The exam is open note. You can use an unlimited number of handwritten cheat sheets, but you must work alone.

We will have a form where you can ask clarification questions. Please check the clarifications page periodically during the exam.

MANDATORY - Honor Code

Read the following honor code and sign your name on your answer sheet. ***Failure to do so will result in a grade of 0 for this exam.***

I understand that I may not collaborate with anyone else on this exam, or cheat in any way. I am aware of the Berkeley Campus Code of Student Conduct and acknowledge that academic misconduct will be reported to the Center for Student Conduct and may further result in partial or complete loss of credit.

Q1 *True/False* (14 points)

Each true/false is worth 2 points unless otherwise specified.

[NOTE: The first question on the exam will always be True/False.]

Q1.1 TRUE or FALSE: Suppose there is a transmission error in a block B of ciphertext using CBC mode. This error propagates to every subsequent block in decryption, which means that the block B and every block after B cannot be decrypted correctly.

○ TRUE ● FALSE

Solution: False. Only B and the block after B are decrypted incorrectly.

Q1.2 TRUE or FALSE: The IV for CBC mode must be kept secret.

○ TRUE ● FALSE

Solution: False. It can be public. For instance, it is normally sent in the clear along with the ciphertext, so any eavesdropper can see the IV—this does not cause any security problems.

Q1.3 TRUE or FALSE: The random number r in El Gamal can be made public.

○ TRUE ● FALSE

Solution: False. If it is public, an attacker can calculate $B^{-r}c_2$ (since B is public).

Q1.4 TRUE or FALSE: If the daily lottery numbers are truly random, then they can be used as the entropy for a one-time-pad since a one-time-pad needs to be random.

○ TRUE ● FALSE

Solution: False, since the information is public.

Q1.5 TRUE or FALSE: It is okay if multiple people use the same modulus p for their El Gamal public key.

● TRUE ○ FALSE

Solution: True, since p is public and known.

Q1.6 TRUE or FALSE: Alice and Bob share a symmetric key k. Alice sends Bob a message encrypted with k stating, "I owe you $100", using AES-CBC encryption. Assuming AES is secure, we can be confident that an active attacker cannot tamper with this message; its integrity is protected.

○ TRUE ● FALSE

Solution: False. An attacker can still modify the ciphertext sent, and there is no way for Bob to tell if the message has been modified.

Q1.7 TRUE or FALSE: Alice and Bob share a secret symmetric key k which they use for calculating MACs. Alice sends the message M = "I, Alice, owe you, Bob, $100" to Bob along with its message authentication code $\text{MAC}_k(M)$. Bob can present $(M, \text{MAC}_k(M))$ to a judge as proof that Alice owes him $100 since a MAC provides integrity.

○ TRUE ● FALSE

Solution: False. A MAC provides integrity, but not does not prove that Alice generated the MAC. Bob can create MACs himself and so that does not prove that Alice wrote the message.

[NOTE: You may not need all blanks on the answer sheet for a question. For this question, you should leave Q1.8-Q1.10 blank.]

This is the end of Q1. Proceed to Q2 on your answer sheet.

Q2 *Hashing Functions* (12 points)

Recall the definition of "one-way functions" and "collision-resistance" from lecture. We say a function f is one-way if given $f(x)$ it is hard to find x' such that $f(x') = f(x)$. Likewise, we say a function f is "collision-resistant" if it is hard to find two inputs x, y such that $f(x) = f(y)$ but $x \neq y$.

For each of the given functions H below, determine if it is one-way or not, and if it is collision-resistant or not.

Q2.1 (3 points) $H(x) = x$

[NOTE: Your answer sheet has six answer choices for every subpart, but not every question will have six answer choices. For example, for Q2.1 here, you should not use options (E) and (F) on your answer sheet.]

○ (A) One-way ○ (C) Both ○ (E) ——

● (B) Collision-resistant ○ (D) Neither ○ (F) ——

Solution: This function is collision-resistant because given two different inputs, $x' \neq x$, the hashes $H(x) = x$ and $H(x') = x'$ are always different.

This function is not one-way because given $H(x)$, we can use it directly as the input to the hash function to get $H(H(x)) = H(x)$.

Q2.2 (3 points) $H(x) = x \bmod 2$

[NOTE: The answer choices on this subpart are circular bubbles, so you should only bubble in one option out of (G), (H), (I), and (J) on your answer sheet for Q2.2.]

○ (G) One-way ○ (I) Both ○ (K) ——

○ (H) Collision-resistant ● (J) Neither ○ (L) ——

Solution: This function is not collision-resistant. Consider $H(0) = H(2) = 0$.

This function is not one-way because given $H(x) = 0$, we know any even value of x will satisfy $H(x) = 0$.

Q2.3 (3 points) $H(x) = E_k(x)$, where where E_k is a ideally secure block cipher with a known and published key k.

○ (A) One-way ○ (C) Both ○ (E) ——

● (B) Collision-resistant ○ (D) Neither ○ (F) ——

Solution: This function is collision-resistant. The output of an ideally secure block cipher is indistinguishable from a random permutation of bits, so it is hard to find two different. inputs that hash to the same output.

This function is not one-way because the key is known and published, so given $H(x)$, we can calculate $D_k(H(x)) = D_k(E_k(x)) = x$, and use this as the input to the hash to get $E_k(x) = H(x)$.

Q2.4 (3 points) $H(x) = 0$

○ (G) One-way ○ (I) Both ○ (K) ——

○ (H) Collision-resistant ● (J) Neither ○ (L) ——

Solution: This function is not collision-resistant. Consider $H(0) = H(1) = 0$.

This function is not one-way because given $H(x) = 0$, we know any value of x will satisfy $H(x) = 0$.

[NOTE: You should leave Q2.5 and Q2.6 blank on your answer sheet. Also, since none of the subparts of this question asked for a short answer, you should leave the space below Q2 on your answer sheet blank.]

This is the end of Q2. Proceed to Q3 on your answer sheet.

Q3 *Finding Common Patients* (35 points)

Caltopia has two hospitals: Bear Hospital and Tree Hospital, each with a database of confidential medical records. Each record is a tuple (p_i, m_i), where p_i is a patient's full name and m_i is the patient's medical record. Each Caltopian citizen has a unique name.

Each hospital has a list of records, $(x_1, m_1), ..., (x_n, m_n)$ for Bear Hospital and $(y_1, m_1), ..., (y_n, m_n)$ for Tree Hospital.

Note: the values of m_i may differ between the two hospitals, even for the same patient.

The two hospitals wish to identify patients that attend both hospitals, but are afraid of eavesdroppers like Eve (who has a list of all the plaintext names of the citizens of Caltopia) listening in.

Bear Hospital and Tree Hospital share a key k that is not known to anyone else. Assume r_i is some random bitstring, and $||$ is the bitwise concatenation operation.

Q3.1 (5 points) Tree Hospital suggests applying some cryptographic function to its list of users, transforming it into a list $(y_1^*, y_2^*, ..., y_n^*)$, which it will send to Bear Hospital.

Bear Hospital will then either decrypt y_i^*, or compute x_i^* in the same way and compare (whichever is appropriate).

Which of the following give y_i^* such that Eve cannot win the IND-CPA game? Select all that apply.

[NOTE: The answer choices on this subpart are square bubbles, so you should bubble in any of options (A), (B), (C), (D), (E) you think are correct, or only option (F) if you think all options are incorrect.]

□ (A) $y_i^* = \text{SHA}(y_i)$

□ (B) $y_i^* = r_i||\text{SHA}(y_i||r_i)$

■ (C) $y_i^* = \text{AES}_k(r_i)||\text{SHA}(y_i||r_i)$

■ (D) $y_i^* = \text{AES-CBC}_k(y_i)$

■ (E) $y_i^* = \text{AES-CBC}_k(\text{SHA}(y_i))$

□ (F) None of the above

Solution: With the first two options, the attacker can test a guess at y_i and determine whether it is correct, since SHA can be computed by anyone; therefore, those options are not IND-CPA-secure. AES encryption of r_i ensures the attacker cannot learn r_i, and that prevents the attacker from learning any information about y_i in the third option. AES-CBC encryption is IND-CPA-secure.

Q3.2 (5 points) For the rest of the question, assume that the lengths of each name are different, and each name is between 1 and 127 bytes long.

Which of the following give y_i^* such that Eve cannot learn anything about the patient names in this new threat model? Select all that apply.

[NOTE: On "select all that apply" questions, you get 1 point for every correct option you choose, and 1 point for every incorrect option that you correctly leave blank (every option is graded independently). For example, if the correct answer is (G) and (H), and you choose (G) and (I), you would get 3/5 points (+1 for choosing (G), and +2 for leaving (J) and (K) blank).]

□ (G) $y_i^* = \text{SHA}(y_i)$

□ (H) $y_i^* = r_i \| \text{SHA}(y_i \| r_i)$

■ (I) $y_i^* = \text{AES}_k(r_i) \| \text{SHA}(y_i \| r_i)$

□ (J) $y_i^* = \text{AES-CBC}_k(y_i)$

■ (K) $y_i^* = \text{AES-CBC}_k(\text{SHA}(y_i))$

□ (L) None of the above

Solution: The first two choices are insecure for the same reasons as in Q3.1. AES-CBC mode is insecure because it reveals partial information about the length of the name; therefore it might be used to infer the name of a patient (e.g., if there is only a single patient whose name is 96-111 bytes long, then that patient can be uniquely identified based on the length of the AES-CBC ciphertexts). The third and fifth option do not reveal any information about the length of the name or the name itself.

Q3.3 (4 points) The hospitals soon realize that, due to privacy laws, they cannot share any plaintext information about patients *even with each other* (including their names) unless **both** hospitals know in advance that a patient has in fact used both hospitals.

Bear Hospital will transform its names using $x_i^* = F_k(x_i)$, and Tree Hospital using $y_i^* = F_k(y_i)$, for some function F. A trusted third party S agrees to take the transformed names $x_1^*, \ldots, x_n^*, y_1^*, \ldots, y_n^*$ from both hospitals, and compute a set of pairs:

$P = \{(i, j) : x_i^* = y_j^*\}$

We want to ensure three requirements with the above scheme:

1. if $x_i = y_j$, then $(i, j) \in P$,
2. if $x_i \neq y_j$, then it is very unlikely that $(i, j) \in P$,
3. even if Eve compromises S, she cannot learn the name of any patient at either hospital or the medical information for any patient.
4. Your solution must still provide guarantee 3 even if a new user with an extremely long (and unique) name were to join Caltopia.

Below are potential candidates for a function F. Select all candidates that meet all four requirements.

□ (A) $F_k(p) = \text{AES-ECB}_k(p)$

□ (B) $F_k(p) = \text{AES-CBC}_k(p)$

□ (C) $F_k(p) = \text{SHA}(p)$

■ (D) $F_k(p) = \text{SHA}(p \| k)$

□ (E) None of the above

□ (F) ——

Solution: Options (A) and (B) reveal partial information about the length of p, which violates condition 4.

Option (C) allows the attacker to test a guess at p (perform a dictionary attack).

Q3.4 (3 points) Same question as the previous part. Select all candidates that meet all four requirements.

[NOTE: Not every question will have six answer choices. For Q3.4, you should only bubble any of options (G), (H), (I) you think are correct, or (J) if you think all options are incorrect. You should leave options (K) and (L) on your answer sheet blank.]

□ (G) $F_k(p) = \text{SHA}(p)||\text{SHA}(k)$ □ (J) None of the above

□ (H) $F_k(p) = \text{AES}_k(r_i)||\text{SHA}(p||r_i)$ □ (K) ——

■ (I) $F_k(p) = \text{AES-ECB}_k(\text{SHA}(p))$ □ (L) ——

Solution: Option (G) allows the attacker to test a guess at p (perform a dictionary attack).

Option (H) violates requirement 1, due to the randomness.

Q3.5 (3 points) Same question as the previous part. Select all candidates that meet all four requirements.

□ (A) $F_k(p) = \text{AES-CBC}_k(\text{SHA}(p))$ □ (D) None of the above

■ (B) $F_k(p) = \text{SHA}(\text{AES-ECB}_k(p))$ □ (E) ——

□ (C) $F_k(p) = \text{SHA}(\text{AES-CBC}_k(p))$ □ (F) ——

Solution: Options (A) and (C) violate requirement 1, due to the random IV in CBC mode.

Q3.6 (5 points) One possible choice for $F_k(p)$ is $\text{SHA}(k||p)$, where $||$ is the concatenation operation.

Explain why $F_k(p) = \text{SHA}(k||p)$ meets requirement (1) (if $x_i = y_j$, then $(i, j) \in P$).

□ (G) —— □ (H) —— □ (I) —— □ (J) —— □ (K) —— □ (L) ——

[NOTE: Since this question has a large box and no multiple-choice options, you should leave the bubbles for Q3.6 blank on your answer sheet, and write your short answer for Q3.6 in the blank space below Q3 on your answer sheet.]

Solution: This is a deterministic function of the name, so if $x_i = y_j$ then $x_i^* = y_j^*$.

Q3.7 (5 points) Explain why $F_k(p) = \text{SHA}(k||p)$ meets requirement (2) (if $x_i \neq y_j$, then it is very unlikely that $(i, j) \in P$).

□ (A) —— □ (B) —— □ (C) —— □ (D) —— □ (E) —— □ (F) ——

[NOTE: If a question has multiple short-answer questions, please be sure to clearly label your answers to each subpart. If you need more space to write your answers, feel free to use another sheet of paper and be sure to upload it with the rest of your answer sheet.]

Solution: SHA-256 is believed to be collision-resistant, so it's hard to find collisions in it. Therefore, it's unlikely that any of the Bear Hospital and Tree Hospital patients' names will cause a collision in SHA-256. If we had $(i, j) \in P$ with $x_i \neq y_j$, that would mean $k||x_i$ has the same SHA-256 hash as $k||y_j$, i.e., we would have found a collision for SHA-256; since SHA-256 is believed to be secure, it's very unlikely this will happen.

Q3.8 (5 points) Explain why $F_k(p) = \text{SHA}(k||p)$ meets requirements (3) and (4).

□ (G) —— □ (H) —— □ (I) —— □ (J) —— □ (K) —— □ (L) ——

Solution: The hash hides the length of each patient's name, and the key prevents the attacker from computing this mapping and testing guesses at the patient names.

[NOTE: The actual exam will have more than 3 questions, but since this is the practice exam, we will stop here. You should now scan or take pictures of your answer sheet and upload it to the "Homework 3" assignment on Gradescope. Thank you for trying out the practice exam!]

This is the end of Q3. You have reached the end of the exam.