CS255: Cryptography and Computer Security **Winter 2018**

Assignment #1

Due: 11:59pm on Mon., Jan. 29, 2018, by Gradescope (each answer on a separate page)

Problem 1. The trouble with compression. Let (E, D) be a semantically secure cipher that operates on messages in $\{0,1\}^{\leq n}$ (i.e. messages whose length is at most n bits). Suppose that the ciphertext output by the encryption algorithm is exactly 128 bits longer than the input plaintext. To reduce ciphertext size, there is a strong desire to combine encryption with lossless compression. We can think of compression as a function from $\{0,1\}^{\leq n}$ to $\{0,1\}^{\leq n}$ where, for some messages, the output is shorter than the input. As always, the compression algorithm is publicly known to everyone.

a. Compress-then-encrypt: Suppose the encryptor compresses the plaintext message m before passing it to the encryption algorithm E. Some n-bit messages compress well, while other messages do not compress at all. Show that the resulting system is not semantically secure by exhibiting a semantic security adversary that obtains advantage close to 1.

b. Encrypt-then-compress: Suppose that instead, the encryptor applies compression to the output of algorithm E (here you may assume the compression algorithm takes messages of length up to $n + 128$ bits as input). Explain why this proposal is of no use for reducing ciphertext size.

Problem 2. The movie industry wants to protect digital content distributed on DVD's. We study one possible approach. Suppose there are at most a total of n DVD players in the world (e.g. $n = 2^{32}$). We view these n players as the leaves of a binary tree of height $\log_2 n$. Every node v_j in this binary tree contains an AES key K_j. These keys are kept secret from consumers and are fixed for all time. At manufacturing time every DVD player is assigned a serial number $i \in [0, n-1]$. Consider the set S_i of $1 + \log_2 n$ nodes along the path from the root to leaf number i in the binary tree. The manufacturer of the DVD player embeds in player number i the $1 + \log_2 n$ keys associated with the nodes in S_i. In this way each DVD player ships with $1 + \log_2 n$ keys embedded in it (these keys are supposedly inaccessible to consumers). A DVD movie M is encrypted as

$$DVD \;=\; \underbrace{E_{K_{root}}(K)}_{\text{header}} \;\Big\|\; \underbrace{E_K(M)}_{\text{body}}$$

where K is some random AES key called a content-key. Since all DVD players have the key K_{root} all players can decrypt the movie M. We refer to $E_{K_{root}}(K)$ as the header and $E_K(M)$ as the body. In what follows the DVD header may contain multiple ciphertexts where each ciphertext is the encryption of the content-key K under some key K_i in the binary tree.

a. Suppose the $1 + \log_2 n$ keys embedded in DVD player number r are exposed by hackers and published on the Internet (say in a program like DeCSS). Show that when the movie industry is about to distribute a new DVD movie they can encrypt the contents of the DVD using a

header of size $\log_2 n$ so that all DVD players can decrypt the movie except for player number r. In effect, the movie industry disables player number r.
Hint: the header will contain $\log_2 n$ ciphertexts where each ciphertext is the encryption of the content-key K under certain $\log_2 n$ keys from the binary tree.

b. Suppose the keys embedded in k DVD players $R = \{r_1, \ldots, r_k\}$ are exposed by hackers. Show that the movie industry can encrypt the contents of a new DVD using a header of size $O(k \log n)$ so that all players can decrypt the movie except for the players in R. You have just shown that all hacked players can be disabled without affecting other consumers.

Side note: the AACS system used to encrypt Blu-ray and HD-DVD disks uses a related system. It was quickly discovered that hackers can expose player secret keys faster than the MPAA can revoke them.

Problem 3. The purpose of this problem is to clarify the concept of *advantage*. Consider the following two experiments EXP(0) and EXP(1):

- In EXP(0) the challenger flips a fair coin (probability 1/2 for HEADS and 1/2 for TAILS) and sends the result to the adversary $\mathcal{A}$.
- In EXP(1) the challenger always sends TAILS to the adversary.

The adversary's goal is to distinguish these two experiments: at the end of each experiment the adversary outputs a bit 0 or 1 for its guess for which experiment it is in. For $b = 0, 1$ let W_b be the event that in experiment b the adversary output 1. The adversary tries to maximize its distinguishing advantage, namely the quantity

$$\text{Adv} = \left| \Pr[W_0] - \Pr[W_1] \right| \quad \in [0, 1] \ .$$

The advantage Adv captures the adversary's ability to distinguish the two experiments. If the advantage is 0 then the adversary behaves exactly the same in both experiments and therefore does not distinguish between them. If the advantage is 1 then the adversary can tell perfectly what experiment it is in. If the advantage is negligible for all efficient adversaries (as defined in class) then we say that the two experiments are indistinguishable.

a. Calculate the advantage of each of the following adversaries:
    - $\mathcal{A}_1$: Always output 1.
    - $\mathcal{A}_2$: Ignore the result reported by the challenger, and randomly output 0 or 1 with even probability.
    - $\mathcal{A}_3$: Output 1 if HEADS was received from the challenger, else output 0.
    - $\mathcal{A}_4$: Output 0 if HEADS was received from the challenger, else output 1.
    - $\mathcal{A}_5$: If HEADS was received, output 1. If TAILS was received, randomly output 0 or 1 with even probability.

b. What is the maximum advantage possible in distinguishing these two experiments? Explain why.

Problem 4. Exercising the definition of semantic security. Let (E, D) be a semantically secure cipher defined over $(\mathcal{K}, \mathcal{M}, \mathcal{C})$, where $\mathcal{M} = \mathcal{C} = \{0, 1\}^L$. Which of the following encryption algorithms yields a semantically secure scheme? Either give an attack or provide a security proof. To prove security, prove the contra-positive, that is prove that a semantic security attacker $\mathcal{B}$ on the proposed system gives a semantic security attacker $\mathcal{A}$ on (E, D), with the same advantage.

a. $E_1(k, m) := 0 \parallel E(k, m)$

b. $E_2(k, m) := E(k, m) \parallel \text{parity}(m)$

c. $E_3(k, m) := \text{reverse}(E(k, m))$

d. $E_4(k, m) := E(k, \text{reverse}(m))$

Here, for a bit string s, parity(s) is 1 if the number of 1's in s is odd, and 0 otherwise; also, reverse(s) is the string obtained by reversing the order of the bits in s, e.g., reverse$(1011) = 1101$.

Problem 5. Let us see why in CBC mode an unpredictable IV is necessary for CPA security.

a. Suppose a defective implementation of CBC encrypts a sequence of packets by always using the last ciphertext block of packet number i as the IV for packet number $i + 1$ (up until a few years ago all web browsers implemented CBC this way). Construct an efficient adversary that wins the CPA game against this implementation with advantage close to 1. Recall that in the CPA game the attacker submits packets (a.k.a messages) to the challenger one by one and receives the encryption of those packets. The attacker then submits the semantic security challenge which the challenger treats as the next packet in the packet stream.

b. Can you suggest a simple fix to the problem from part (a) that does not add any additional bits to the ciphertext?

c. Suppose the block cipher (E, D) used for CBC encryption has a block size of n bits. Construct an attacker that wins the CPA game against CBC with a random IV (i.e. where the IV for each message is chosen independently at random) with advantage close to $1/2^n$.

Your answer for part (c) explains why CBC cannot be used with a block cipher that has a small block size (e.g. $n = 32$ bits). Note that there are many other problems with such a small block size, which is why AES has a block size of 128 bits.

Problem 6. Let F be a secure PRF defined over $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$, where $\mathcal{K} = \mathcal{X} = \mathcal{Y} = \{0, 1\}^n$.

a. Show that $F_1\big(k,\ (x_1, x_2)\big) := F(k, x_1) \oplus F(k, x_2)$ is insecure. That is show an attacker $\mathcal{A}_1$ on F_1 that has non-negligible advantage in distinguishing $F_1(k, \cdot)$ from a random function.

b. Prove that $F_2(k, x) := F(k, x) \oplus x$ is a secure PRF. Do so using the contra-positive: show that if an adversary $\mathcal{A}_2$ can distinguish $F_2(k, \cdot)$ from a random function then there is adversary $\mathcal{B}$ (that is a wrapper around $\mathcal{A}_2$) that can distinguish F from a random function.

Problem 7. Let $\mathcal{E} = (E, D)$ be a cipher. Consider the cipher $\mathcal{E}_2 = (E_2, D_2)$, where $E_2(k, m) = E(k, E(k, m))$. One would expect that if encrypting a message once with E is secure then encrypting it twice as in E_2 should be no less secure. However, that is not always true.

a. Show that there is a semantically secure cipher $\mathcal{E}$ such that $\mathcal{E}_2$ is not semantically secure.

b. Prove that for every CPA secure ciphers $\mathcal{E}$, the cipher $\mathcal{E}_2$ is also CPA secure. That is, show that for every CPA adversary $\mathcal{A}$ attacking $\mathcal{E}_2$ there is a CPA adversary $\mathcal{B}$ (that uses $\mathcal{A}$ as a black box) attacking $\mathcal{E}$ with about the same advantage and running time.