Total points: 80 (+6 bonus)

Assignment 2

Due: 11:59 PM, 16th July (Fri)

Written assignment

1. Bob runs a social network web server that requires users, such as Alice, to log in. He wants to use cryptography to protect Alice's account. However, he doesn't understand cryptography very well. For each of the below uses of cryptography, identify his mistake, and explain what he should do instead.

   (a) [4 points] Bob obtains a public/private encryption key pair using RSA. When Alice visits his site, Alice generates a 128-bit secret key intended for AES in CBC mode, encrypts it using Bob's RSA public key, and sends it to Bob. Bob asks Alice to use a SHA-256 HMAC based on the same 128-bit secret key to authenticate this message so as to prove that it is indeed Alice who sent the key.

   (b) [4 points] To establish trust, Bob asks a CA to sign his public 256-bit RSA key using the CA's private 256-bit ECC key. The CA's public 256-bit ECC key is in Alice's browser, so Alice can verify Bob's key automatically when she visits his site, even though she does not explicitly know who the CA is.

   (c) [4 points] For Alice's login, Bob requires Alice to hash and salt her password on the client side using SHA-512, and then send it to Bob using 128-bit AES. Then, Bob will store Alice's hashed password and the salt in his database. In future attempts, Alice can use the same hashed password and salt to login.

   (d) [4 points] To store Alice's password securely, Bob uses AES encryption with a secret key and a 128-bit IV to encrypt her password. A CRC32 checksum is used to ensure correctness against random bit flip errors.

2. [9 points] When a client C accesses server S through Tor, she usually builds a circuit of three nodes: N_1, N_2, and N_3. A connection is established as follows:

$$C \to N_1 \to N_2 \to N_3 \to S$$

N_1 is also known as the entry node or the guard node, and N_3 is also known as the exit node. Visit `metrics.torproject.org` to answer the following questions:

(a) [3 points] Give the total amount of advertised bandwidth of relays with the "Guard" flag (but not the "Exit" flag) and relays with the "Exit" flag (but not the "Guard" flag) on 2020-02-01. Which is more? Give one reason to explain this phenomenon.

(b) [3 points] Give the median download rate of a file (in bits per second) for a 50 KiB file and a 5 MiB file to the `op-hk` onion server on 2020-02-01. Can you explain the difference?

(c) [3 points] What is a disadvantage of using three nodes in a Tor circuit instead of one node? What is an advantage of doing so?

Programming assignment

Breaking Cryptography

In this assignment, we will write programs to automatically break some weak ciphers. Please make sure to read the submission instructions carefully.

Two-Time Pad [25 points]

Two files, `ctext0` and `ctext1`, have been sent to you by e-mail. Those two files were encrypted using the same one-time pad. They are exactly 400 bytes each, and they both come from popular English Wikipedia articles. Find the contents of both files using crib-dragging, and submit them as `ptext0` and `ptext1`. You can flip around ptext0 and ptext1.

You may assume the plaintext to consist only of ASCII characters with the following byte values, all ranges being inclusive of both ends:

- Symbols: 32 to 41, 44 to 59, 63, 91, 93.
- Capital letters: 65 to 90.
- Small letters: 97 to 122.

The ctext file was derived by XOR'ing the plaintext (as ASCII bytes) with the 400-byte key. Each byte of the ctext file would be the XOR of the corresponding byte of plaintext with the corresponding byte of the key (written as bit strings). If you cannot find the full texts, submit as much of the text as you can find.

Padding Oracle Attack [30 points]

AES — the standard block cipher in use today — had a padding algorithm that introduced vulnerabilities when combined with CBC (Ciphertext Block Chaining). In this assignment, we will investigate why it was insecure. In fact, the attacker can arbitrarily decrypt and encrypt in AES without knowledge of the key, and even without any understanding of the operations of AES.

The following is a partial adaptation of Vaudenay's "Security Flaws Induced by CBC Padding — Applications to SSL, IPSEC, WTLS ..." paper. You may skip the explanation here and read the first four pages of that paper to answer the questions directly. Note that AES operates on blocks of 16 bytes instead of 8 in that paper.

AES encrypts plaintexts in blocks of 16 bytes at a time. If there are fewer than 16 bytes of plaintext data, AES adds **padding** bytes to the end of the plaintext until there are 16 bytes exactly. (During decryption, those padding bytes will be discarded.) If there are more than 16 bytes of data, AES operates on each block one by one in order, and pads the final block to 16 bytes. If we need to add n bytes of padding, then the bytes to add is exactly n copies of n. For example, suppose the plaintext we want to encrypt is:

$$x' = (\texttt{CA013AB4C561})_{16}$$

In the above, x' is written in hexadecimal notation, and it has 6 bytes. We want to add 10 bytes to make 16 bytes, so we will add the byte $(\texttt{0A})_{16} = 10$ ten times to make x, the padded version of x':

$$x = (\texttt{CA013AB4C5610A0A0A0A0A0A0A0A0A0A})_{16}$$

Note that the minimum amount of padding is 1 byte: that is to say, if the original plaintext has a multiple of 16 bytes, then we will need to add 16 bytes of padding of $(\texttt{10})_{16} = 16$. There will be a whole block of padding at the end.

After padding x' to x, we can perform AES encryption (denote the operation as C) on x to get the ciphertext $C(x)$. C is dependent on the secret key K and the initialization vector IV; the attacker knows IV because it is sent in the clear.

Suppose x contains N blocks of data (in other words, the size of x is $16N$ bytes), denoted as $(x_1|x_2|\ldots|x_N)$. $|$ is the concatenation operation, meaning that the bytes of x_1 are followed by that of x_2, and then by x_3, and so on. After encryption, the resulting ciphertext is $(IV|y_1|y_2|\ldots|y_N)$. In CBC mode, we have:

$$\begin{aligned} y_1 &= C(IV \oplus x_1) \\ y_i &= C(y_{i-1} \oplus x_i) \text{ for } i = 2, 3, \ldots, N \end{aligned}$$

The inverse of C, the AES block encryption function, is denoted as D, the block decryption function. Note that both C and D do not perform any padding on their own; they both input and output 16 bytes of data. For any 16-byte block z, $D(C(z)) = z$.

We will now break AES in CBC mode using a *padding oracle*. A padding oracle is some entity that tells the attacker if the padding of some ciphertext $(IV|y_1|\ldots|y_N)$ is correct after decryption. In other words, it decrypts $(IV|y)$ using the correct key, gets the plaintext x, and checks if x uses the correct padding scheme described above. The padding oracle has been shared with you. (See "Notes on the Padding Oracle" later for more details on how to run the padding oracle.)

Suppose we are deciphering some ciphertext $(IV|y_1|\ldots|y_N)$. There will be three steps. First, we will learn how to find the last byte of x_N ("Decrypt byte"). Then, we will find the whole x_N ("Decrypt block"). Finally, we will find all of $(x_1|x_2|\ldots|x_N)$ ("Decrypt").

— Decrypt byte —

Extract y_N from the ciphertext by taking the last 16 bytes, and y_{N-1} as the last 32 to 16 bytes. Denote the ith byte of y_N as $y_{N,i}$. Here, we want to find $x_{N,16}$.

1. First, generate a random block $r = (r_1|r_2|\ldots|r_{15}|i)$ with 15 random bytes, followed by a byte i. Initially $i = 0$.
2. Ask the padding oracle if $(r|y_N)$ is valid. $(r|y_N)$ contains the 16 bytes of r, followed by the 16 bytes of y.
3. If the padding oracle returns "no", increment i by 1, and then ask the padding oracle again. Keep incrementing i until the padding oracle returns "yes".

4. Replace r_1 with any other byte and ask the oracle if the new $(r|y_N)$ has valid padding. If the padding oracle returns "yes", similarly replace r_2. Repeat until either we have finished replacing r_{15} and the oracle always returned "yes", or the oracle has returns "no" while we were replacing some r_k.
5. If the oracle always returned "yes" in Step 4, set $D(y_N)_{16} = i \oplus 1$.
6. If the oracle returned "no" when we replaced r_k in Step 4, set $D(y_N)_{16} = i \oplus (17-k)$.
7. The final byte of x_N is $x_{N,16} = D(y_N)_{16} \oplus y_{N-1,16}$.

— *Decrypt block* —

After finding $x_{N,16}$, the attacker can proceed to find all other bytes of x_N, starting from the 15th byte $x_{N,15}$, then $x_{N,14}$, and proceeding backwards to $x_{N,1}$. In this process, the attacker will also find $D(y_N)_{16}, D(y_N)_{15}, \ldots, D(y_N)_1$ as above. The following describes how the attacker can find $x_{N,k}$ for any k; the attacker has already found $D(y_N)_{k+1}, D(y_N)_{k+2}, \ldots, D(y_N)_{16}$.

1. Set r as $(r_1|r_2|\ldots|r_{k-1}|i|D(y)_{k+1} \oplus (17-k)|D(y)_{k+2} \oplus (17-k)|\ldots|D(y)_{16} \oplus (17-k))$. Initially $i = 0$.
2. Ask the oracle if $r|y_N$ is valid.
3. If the padding oracle returns "no", increment i and ask the padding oracle again. Keep incrementing i until the padding oracle returns "yes".
4. When the padding oracle returns "yes", set $D(y_N)_k = i \oplus (17 - k)$
5. The k-th byte of x_N is $x_{N,k} = D(y_N)_k \oplus y_{N-1,k}$.

— *Decrypt* —

The above shows how the attacker can decrypt the last block y_N to obtain X_N. To decrypt the k-th block y_k, the attacker simply replaces all of the above y_N with y_k and y_{N-1} with y_{k-1}.

Your task is to write a program, `decrypt`, which finds the plaintext x for any ciphertext y and outputs it to standard output. It is run with:

```
./decrypt ciphertext
```

`ciphertext` is a file that contains an amount of data that is a multiple of 16 bytes, and at least 32 bytes. It is formatted as $IV|y_1|\ldots|y_N$, where the IV is the first 16 bytes, y_1 are bytes 17 to 32, and so on.

After you get the plaintext, output it to standard output. Do not add a newline.

This is a difficult task. You should tackle the assignment step by step: do the "Decrypt byte" step, then the "Decrypt block" step, then the "Decrypt" step. In case you cannot finish the assignment, marks will be given for partially completing each step.

Hint: Suppose you are given the ciphertext $(IV|y_1|y_2)$. Write down the plaintext $(x_1|x_2)$ using D, IV, y_1, and y_2. (It is not simply $D(y_1)$ and $D(y_2)$.)

Bonus (6 points) Write a program, `encrypt`, which takes in some plaintext x and encrypts x using the same encryption algorithm and key that is behind the padding oracle provided. It is run with:

```
./encrypt plaintext
```

`plaintext` contains an amount of data that is a multiple of 16 bytes, and at least 16 bytes. It is formatted as $x_1|x_2|\ldots|x_N$. Output the ciphertext and the IV to standard output as $IV|y_1|\ldots|y_N$.

(Hint: `encrypt` should call `decrypt` as a subroutine in order to guess the right ciphertext. You only need to call `decrypt` once for each block. Note that you can choose your own IV.)

Notes on the padding oracle

The padding oracle should be run with:

```
python3 oracle.py ciphertext
```

It will decrypt the ciphertext with the secret AES key, check the padding of the plaintext, and output "1" if the padding is correct and "0" if the padding is incorrect.

The padding oracle was written in Python. It is not compiled, and it can be directly run on Unix-like systems such as Ubuntu and macOS. If you want to run it on Windows, you will have to install Python and then type:

```
python3 oracle.py ciphertext
```

You will also have to capture the output and feed it into your own code. The command to do so is `system(<your command>)` in C and C++, `subprocess.check_output(<your command>)` in Python, and `Runtime.getRuntime().exec(<your command>)` in Java. You may have to read about your preferred function specifically to learn how to use it.

Since the oracle is not compiled, the key is hardcoded into the oracle code. **Do not use the key in any way.** When we test your code, we will use a different oracle with a different key. Your code should work independent of what the actual key value is.

You are also provided with a ciphertext called `ciphertext` for reference, with its generator `ciphertext_gen.py`. It was encrypted with the same key as the oracle, with an $IV =$ `CMPT 479 test iv`, and the plaintext message is `Alice and Bob in Wonderland`. Since the plaintext is 27 bytes long, it will be padded with 5 bytes, each with a byte value of 5. If you find that the byte value of the last byte of the plaintext is 5, you are on the right track!

Submission instructions

All submissions should be done through CourSys. Submit the following files:

- `a2.pdf`, containing all your written answers. Make sure it is not the question file.
- `ptext0` and `ptext1`, for part (a) of the programming assignment.
- `decrypt.{cpp, py, java}`, for part (b) of the programming assignment, as well as any other code necessary to run it. This may include a Makefile. Submit your code; do not submit any compiled files. You may also submit `encrypt.{cpp, py, java}` for the bonus marks. The bonus marks can only be applied to this assignment.

To run decrypt, for example, I will do the following:

C++: I will compile `./gcc decrypt.cpp -o decrypt` and then run `./decrypt`.

Python3: I will call `python decrypt.py`.

Java: I will compile `javac decrypt.java` and then call `java decrypt`.

If you are using Python, pleaase make sure it is Python3 instead of Python2.

If there is a Makefile in your folder, the Makefile will override all of the above. I will call `make` to compile the code, and then I will call `make run`.

Keep in mind that plagiarism is a serious academic offense; you may discuss the assignment, but write your assignment alone and do not show anyone your answers and code.

The submission system will be closed exactly 48 hours after the due date of the assignment. You will receive no marks if there is no submission within 48 hours after the due date.