CS 6324: Information Security **Spring 2021**

Project #1 - Encrypted File System

Zhen Wang, zxw180035

Due date & time: 11:59pm CST on March 26, 2021. Submit to eLearning by the due time.

Additional Instructions: (1) The submitted homework must be typed. Using Latex is recommended, but not required.

1 Overview

In a traditional file system, files are usually stored on disks unencrypted. When the disks are stolen by someone, contents of those files can be easily recovered by the malicious people.

Encrypted File System (EFS) is developed to prevent such leakages. In an EFS, files on disks are all encrypted, nobody can decrypt the files without knowing the required secret. Therefore, even if a EFS disk is stolen, or if otherwise an adversary can read the file stored on the disk, its files are kept confidential. EFS has been implemented in a number of operating systems, such as Solaris, Windows NT, and Linux.

In this project, you are asked to implement a simulated version of EFS in Java. More specifically, you will need to implement several library functions, which simulates the functionalities of EFS.

1.1 Functionality Requirements

To simulate a file system, files are stored in blocks of fixed size. More specifically, a file will be stored in a directory which has the same name as the file. The file will be split into trunks and stored into different physical files. That is, a file named "/abc.txt" is stored in a directory "/abc.txt/", which includes one or more physical files: "/abc.txt/0", "/abc.text/1", and so on, each physical file is of size exactly 1024 bytes, to simulate a disk block. You will need to implement the following functions:

create(file_name, user_name, password) Create a file that can be opened by someone who knows both user_name and password. Both user_name and password are ASCII strings of at most 128 bytes.

findUser(file_name) Return the user_name associated with the file.

length(file_name, password) Return the length of the file, provided that the given password matches the one specified in creation. If it does not match, your code should throw an exception.

read(file_name, starting_position, length, password) Return the content of the file for the specified segment, provided that the given password matches. If it does not match, your code should throw an exception.

write(file_name, starting_position, content, password) Write content into the file at the specified position, provided that the password matches. If it does not match, the file should not be changed and your code should throw an exception.

cut(file_name, length, password) Cut the file to be the specified length, provided that the password matches. If it does not match, no change should occur and your code should throw an exception.

check_integrity(file_name, password) Check that the file has not been modified outside this interface. If someone has modified the file content or meta-data without using the write call, return False; otherwise, return True, provided that the given password matches. If it does not match, your code should throw an exception.

1.2 Security Requirements

Meta-data storage You will need to store some meta-data for each file. In file systems, this is naturally stored in the i-node data structure. In this project, we require that meta-data are stored as part of the physical files. You will need to decide where to put such data (e.g. at the beginning of the physical files), and also what cryptographic operations need to be performed to the meta-data. Naturally the first physical file would contain some meta-data, however, you can also store meta-data in other physical files.

User Authentication You need to ensure that if the password does not match, reading and writing is not allowed. You thus need to store something that is derived from the password; however, you should make it as difficult as possible for an adversary who attempts to recover the password from the stored information (perhaps using a dictionary attack).

Encryption keys In an EFS, we can choose to use one single key to encrypt all the files in the encrypted file system; or we can choose to encrypt each file using a different key. In this lab, we choose the latter approach, in order to reduce the amount of data encrypted under one key.

Algorithm Choice You are required to use AES with 128-bit block size and 128-bit key size. The code for encrypting/decrypting one block (i.e. 128 bits) is provided. When you encrypt/decrypt data that are more than one blocks, you are required to use CTR, the Counter mode. You will need to decide how to generate the IV's (initial vectors). For encryption, you can treat a file as a message, or treat a chunk (stored in one physical file) as a message, or choose some other design.

We assume that an adversary may read the content of the file stored on disk from time to time. In particular, a file may be written multiple times, and the adversary may observed the content on disk between modifications. Your design and implementation should be secure against such an adversary.

You design should also hide the length of the file as much as possible. The number of physical files used for a file will leak some information about the file length; however, your design should not leak any additional information. That is, if files of length 1700 bytes and 1800 bytes both need 2 physical files, then an adversary should not be able to tell which is the case.

Message Authentication We want to detect unauthorized modification to the encrypted files. In particular, if the adversaries modify the file by directly accessing the disk containing EFS, we want to detect such modifications. Modification to other meta-data such as the user or file length should also be detected. Message Authentication Code (MAC) can help. You need to decide what specific algorithm to use, and how to combine encryption and MAC.

1.3 Efficiency Requirements

We also have the following two efficiency requirements.

Storage We want to minimize the number of physical files used for each file.

Speed We want minimize the number of physical files accessed for each read or write operation. That is, if an write operation changes only one byte in the file, we want to access as small a number of physical files as possible, even if the file is very long.

These two efficiency goals may be mutually conflicting. You need to choose a design that offers a balanced tradeoff.

2 Project Tasks

You are asked to complete "EFS.java". More specifically, you will need to implement the following functions. The functionality of the functions are described in Sec 1.1. Please do NOT modify the other files. If there is indeed a need to modify them, please describe in detail in your report.

You are not allowed to use encryption/decryption/hash/MAC functions in Java library. However, you can use the functions provided by us.

1. create(String file_nameString user_name, String password)
2. String findUser(String file_name)
3. int length(String file_name, String password)
4. byte[] read(String file_name, int starting_position, int length, String password)
5. void write(String file_name, int starting_position, byte[] content, String password)
6. void cut(String file_name, int length, String password)
7. boolean check_integrity(String file_name, String password)

In your implementation, if the password is incorrect, please throw **PasswordIncorrectException**. For other exceptions/errors, you can either throw an exception(other than PasswordIncorrectException), or handle it yourself.

3 Report

In your report, you should include the following:

3.1 Design explanation

meta-data design Describe the precise meta-data structure you store and where. For example, you should describe in which physical file and which location (e.g., in "/abc.txt/0", bytes from 0 to 127 stores the user_name, bytes from 128 to ... stores ...)

I store the meta-data in the first physical file. In meta-data, there are: username, salt, hash(password || salt), IV(counter), encrypted file length, MAC.

user authentication Describe how you store password-related information and conduct user authentication.

I store the salt and hash(password || salt) in the metadata. Once the password is entered, we check if it matches hash(password || salt) and thus ensure authentication.

encryption design Describe how files are encrypted. How files are divided up, and how encryption is performed? Why your design ensures security even though the adversary can read each stored version on disk.

I use AES CTR mode to encrypt the file. It uses a random IV/counter. CTR mode turns a block cipher into a stream cipher. It generates the next keystream block by encrypting successive values of a "counter". For er very succesive AES block, the counter increases by one. The counter serves as the AES message and it will be AESed using the key derived from hashing password. The AES encrypted cipher will be XORed with each block of 16 bytes message. The adversary cannot read the file because he does not know the file text is encrypted, even though he knows the procedure, he cannot break it because he does not have AES key which is not stored but derived from user password.

length hiding Describe how your design hide the file length to the degree that is feasible without increasing the number of physical files needed.

I use AES to encrypt the file length and put the encrypted file length into meta-data. The AES key is not stored but derived by hashing password. By that way, no ineasing the number of physical files needed. I pad 0(s) before the key length, so when decrypt I can get the length by simply discard the left side 0(s).

message authentication Describe how you implement message authentication, and in particular, how this interacts with encryption.

We HMAC password and file. The HMACed value is stored inside metadata, when metadata has been modified without using write(), then we check the integrity and it will return a false because MAV value will not be updated when the attacker tamper with the data. Only the conditions 1) correct password and 2) unmodified file both hold that a true can be returned. For simplicity and the answer from MS Teams, only the integrity of mete-data file is protected.

efficiency Describe your design, and analyze the storage and speed efficiency of your design. Describe a design that offers maximum storage efficiency. Describe a design that offers maximum speed efficiency. Explain why you chose your particular design.

The metadata is stored in the first file which contains username, salt, encrypted hash, IV and HMAC. They are stored via create function. The file is encrypted via AES-CTR mode which is safer than AES-ECB mode. The length, read and write and cut functions are password required only username can be gotten witrhout needing password. By that way, we achieve strong security. The design is pretty efficient becasue we do not conduct heavy computation compared with sample mode and does not require more space for storage the meta data: it still can be kept in only one sub-file, the text content also does not require more space.

3.2 Pseudo-code

Provide pseudo-code for the functions **create, length, read, write, check_integrity, cut**. From your description, any crypto detail should be clear. Basically, it should be possible to check whether your implementation is correct without referring to your source code. You may want to describe how password is checked separately since it is used by several of the functions.

```
password_check(filename,password) {
    read salt and hash1;
```

```
    calculate hash2;
    if (hash1 == hash2)
        password correct
}
```

```
create(){
    create salt
    create IV
    create encrypted hash
    create encrypted 0, length inital value
    create HMAC of password, and metadata
    write the metadata and HMAC into first file
}
```

```
length(){
 password check();
 if correct
    read from metadata
    decrypt length using IV and derived AES key
}
```

```
read(){
 password check();
 if correct
    read whole content of that block(s) using position and size
    decrypt total block(s)
    read from the specified position and size
}
```

```
write(){
 password check();
 if correct
    read whole content of that block(s) using position and size
    decrypt the content
    write new content
    encrypte and write back to the block(s)
    compute new HMAC and update metadata
}
```

```
check_integrity(){
 password check();
 if correct
    read stored MAC from first file
    read metadata and calculate MAC2
    if(MAC==MAC2) not modified
    else metadata has been tampered.
}
```

```
cut(){
  password check();
  if correct
    read block(s) according to the remaining length
    clear files
    write the remaining back to files
    update metadata and HMAC
}
```

3.3 Design variations (10 pts)

1. Suppose that the only write operation that could occur is to append at the end of the file. How would you change your design to achieve the best efficiency (storage and speed) without affecting security?

   Now since I dont need to consider inserting. I do not need to read and decrypt and then insert. I just encrypt the contents and attach it to the end of current blocks.

2. Suppose that we are concerned only with adversaries that steal the disks. That is, the adversary can read only one version of the the same file. How would you change your design to achieve the best efficiency?

   Since the adversary cannot manipulate the input and get the corresponding output. Most attacks will not be possible. For example, we can use ECB mode to encrypt a file and it will has the speed and space benefits.

3. Can you use the CBC mode? If yes, how would your design change, and analyze the efficiency of the resulting design. If no, why?

   CBC mode is not appropriate, because an adversary may use chosen plaintext attack. He write multiple times and keeps the previous part of the content to write the same, by observing the chain, it is possible to break the key.

4. Can you use the ECB mode? If yes, how would your design change, and analyze the efficiency of the resulting design. If no, why?

   ECB mode is insecure. Because using the same key for every block, the chosen plaintext attack would be even easier to implement.

3.4 Paper Reading

It is suggested that you complete the following early, because doing so may help you think about your design.

Problem 1 (10 pts) Read the following paper and write a summary of what you have learned from it.

- Ross Anderson: Why Cryptosystems Fail. (`wcf.pdf` is attached in eLearning.)

This paper discusses why the the present cryptosystem fails in not merely a technical but in a practical view. The discussion is based on a retail banking system example. It incorporates all the possible practical attack implementations:

1) Inside Bank, e.g., bad bank clerk and teller

2) Outside ATM, e.g., PIN leakage.

3) Management and operation issues. e.g., the bank does not keep the PIN confidential

3) Insecure ATM encryption design

4) Problems with encryption products. e.g., bad programmer and technical staff

5) No system level management for both equipment vendors and banks; ramshackle administration.

The takeaway form this paper is that the complicated technical attack we used to focus on is indeed not the key threat of our current cryptosystems. For example, cryptanalysis may be one of the less likely threats to banking systems. Rather, how do we store and manage the key is more important than cryptanalysis.

In other words, Designers of cryptographic systems have suffered from a lack of information about how their products fail in practice, as opposed to how they might fail in theory. This lack of feedback has led to a false threat model being accepted. Designers focused on what could possibly go wrong, rather than on what was likely to; and many of their products are so complex and tricky to use that they are rarely used properly.

To implement a good cryptosystem, at least we need to :

1) List all possible failure modes of the system and corresponding strategies.

2) List all the failure management strategies implementation, including the consequences.

3) The design and usage rules should keep those failures in mind.

3) For both vendors and users of the cryptosystem, pay attention to the system level management and scheduling.

4) Take the procedures and employee training and seriously. A cavalier attitude should never be allowed.

To conclude, the defence and design should be more practical and should be paid great attention to.

Problem 2 (10 pts) Read the following paper and write a summary of what you have learned from it.

- Nikita Borisov, Ian Goldberg, David Wagner: Intercepting Mobile Communications: The Insecurity of 802.11. (`mobicom.pdf` is attached in eLearning.)

This paper discusses the security flaws in WEP protocol (a protocol used to protect link-layer communications from eavesdropping and other attacks) stemming from misapplication of cryptographic primitives. The flaws makes eavesdropping and tampering on the channel possible. The protocol is a symmetric key protocol: relies on shared secret k shared by the sender and receiver, and RC4 is used for creating key stream. The receiver verifies the checksum to check integrity of the message after decryption. The protocol goal is to to protect confidentiality of user data from eavesdropping on the wireless channel; integrity; and access control. But there are several attacks that WEP is vulnerable to.

WEP may leak the messages because it reuses keystreams. The plaintexts can be recovered. Though WEP uses IV to generate a new keystream each time, WEP did not totally avoid this IV repeat Problem. because they are bound to be reused at some point, after thousands of packets., thus keystreams repeat.

Dictionary Attack is also possible. a attacker now can decrypt any message on the network which has the same IV. Over time the attack can creat a dictionary if a mapping of keystream to IV is done. The key can be recovered in fairly less time.

The protocol also suffers from key Management Problem, because it uses single key for whole network, not per client. The more it is used within the network, the IV is more likely to repeat. The checksum is also not secure enough to verify the integrity of messages because checksum is linear

Attacks on Access Control also can happen: the attacker can inject new packets in the network using the knowledge of keystream and IV; it is also possible to reuse old IV values without any alarms being triggered, the attacker can use it to send as many messages in the network.

The attacker can also trick the access point to decrypt the messages using IP redirection and reaction attacks. To against it, we can use VPNs to restrict routes and create firewalls or let every client have its own key.

The takeaway is a perfectly secure protocol is not easy to implement. CRC-32 and RC4 are known for speed and ease of use but have serious flaws in implementation in WEP. Any protols should be examined by the public. Also, WEP has many flaws and fails to protect link-layer communications.

4 Submission and Grading

Please submit your "EFS.java" along with your report.

If you have modified other files, please submit all the source files.

You will be graded on the following criteria:

Functionality Correctness (20 pts) Your code should implement the read/write/cut etc., correctly, so that a sequence of calls will return correct results.

Design Correctness (40 pts) To what degree your design/implementation satisfy the specifications? Are there security vulnerabilities? Etc.

Clarity of report (10 pts) To what extent you report describes the design clearly.

Design variants (10 pts)

Paper reading (20 pts)

5 Project Instruction

You will find all the source code in Project1/src.

We provide a text editor which can be used to verify your design/implementation. To use the editor, simply launch "Editor.java". We also provide a sample program named "Sample.java". You are encouraged to start by reading the file. Please notice that in the sample, files are NOT encrypted. Also, there is NO efficiency optimization. You will need to design them yourself.

Please feel free to use your own program to verify your implementation if you want.

Here are the details of the sample program.

create(file_name, user_name, password) The detailed steps are

1. Create a directory named as the name of the file you want to create (e.g. /abc.txt/).
2. Create the first block and use the whole block as meta-data (/abc.txt/0).
3. Write "0" into the first block. It means the file length is 0.
4. Write user_name into the first block.

findUser(file_name) Find the user_name from the first block (/abc.txt/0).

length(file_name, password) Find the length of file from the first block (/abc.txt/0).

read(file_name, starting_position, length, password) The detailed steps are

1. Compute the block that contains the start position. Assume it is n_1.
2. Compute the block that contains the end position. Assume it is n_2.
3. Sequentially read in the blocks from n_1 to n_2 (/abc.txt/n_1, /abc.txt/(n_1 + 1), ... , /abc.txt/n_2).
4. Get the desired string.

write(file_name, starting_position, content, password) The detailed steps are

1. Compute the block that contains the start position. Assume it is n_1
2. Compute the block that contains the end position. Assume it is n_2.
3. Sequentially write into the blocks from n_1 to n_2 (/abc.txt/n_1, /abc.txt/(n_1 + 1), ... , /abc.txt/n_2).
4. Update the first block(meta data) for the length of file if needed.

cut(file_name, length, password) The detailed steps are

1. Find the number of block needed. Assume it is n.
2. Envoke $write$ to update block n (/abc.txt/$(n + 1)$).
3. Remove the redundant blocks if necessary.
4. Update the first block(meta data) for the length of file.

check_integrity(file_name, password) N/A

We also provide some utility functions in "Utility.java". Since the class "EFS" is derived from the class "Utility". You can directly invoke them. The functions and parameters are listed here. Using these functions is encouraged but not required.

void set_username_password() You can set/reset user name and password by invoking this function.

byte[] read_from_file(File file) The function will return all the content of $file$ as a byte array. So $file$ can be a binary file. It will throw an exception if the file cannot be read.

void save_to_file(byte[] s, File file) The function will write s to a binary file $file$(overwrite). It will throw an exception if the file cannot be written.

File set_dir() It will allow you to select a directory. The return value is the chosen directory. You can choose a new directory by typing the full path. If nothing is chosen, return null.

byte[] encrypt_AES(byte[] plainText, byte[] key) The function will return AES encryption result of $plainText$ using key as key.

byte[] decript_AES(byte[] cipherText, byte[] key) The function will return AES decryption result of $cipherText$ using key as key.

byte[] hash_SHA256(byte[] message) The function will return SHA256 result of $message$.

byte[] hash_SHA384(byte[] message) The function will return SHA384 result of $message$.

byte[] hash_SHA512(byte[] message) The function will return SHA512 result of $message$.

byte[] secureRandomNumber(int randomNumberLength) The function will return a random number vector with length $randomNumberLength$. You can assume this is a secure random number generator.