CS5435: Security and Privacy concepts in the wild

Fall2014

Homework Assignment 1 Write-Up

Jai Chaudhary

jc2855@cornell.edu

Problem 1

Name three incentives an attacker might have for paying $325 (or more) for a high-profile Twitter account password. At least one of these should make money for the attacker. [15 points]

- Bad Publicity of an individual– Groups like Anonymous used twitter feeds for defamation. For example in a cyber-attack against Aaron Barr, CEO of the computer security firm HBGary Federal in retaliation for his research on Anonymous and his threat to expose members of the group.
- Access to other accounts linked to the twitter account for authentication
- Hoax - A hacked Twitter account of a major news organization momentarily sent stocks into free fall on April 24, 2013. A tweet from the Associated Press exclaimed: "Breaking: Two Explosions in the White House and Barack Obama is injured." Within seconds, Wall Street was in panic mode and the Dow Jones industrial average and other benchmark indexes plummeted.

Problem 2

Question 2.1 -For an online service whose users are U.S. inhabitants, give an estimate of the average number of randomly selected accounts an attacker would have to try to answer the question “What’s your mother’s maiden name?” correctly. Assume one guess per account.

- Most common surname is “Smith” of around 2,376,206 people in US having the same surname (http://www.census.gov/genealogy/www/data/2000surnames/index.html).
- Given that the population of America is around 313.9 million, we get .75 % of the people in US have the surname “Smith”, which is approx. one in 134 people.
- So the average number of tries of different accounts that would eventually crack the password would be 134 tries.

QUESTION 2.2: FIND THE WEAKEST PASSWORD-RECOVERY QUESTION YOU CAN ONLINE THAT WASN’T DISCUSSED IN CLASS. GIVE AN ESTIMATE OF THE GUESSING PROBABILITY AND EXPLAIN HOW YOU ARRIVED AT IT. [12 POINTS]

- WHICH IS YOUR FAVORITE SPORTS TEAM?
- There are approximately 1.2 million Manchester united fans present in England
- Population of England is 53 million
- GP IS 0.022

Question 3.1: What adversarial model is the RNB assuming, i.e., what is the RNB assuming about the confidentiality provided by the encryption scheme? [4 points]
Assumption is that the adversary has to guess the key in order to get the pin. The average number of tries to crack the AES-256 key to decrypt the pins will be 2n/2 ie 2 ^128 which is very computational intensive hence almost impossible to crack

Question 3.2: Show an attack outside this model. What PIN belongs to the customer Ebenezer Scrooge? Note that a '*' in the customer name column denotes a customer other than Ebenezer Scrooge. (Hint: See http://www.datagenetics.com/blog/september32012/.) [10 points]

An attack is to see the frequency of a particular pin and match it with a pin that is known which has similar frequency in similar number of items/pins in a set. 184 is the total number of times the ciphertext of Ebenezer Scrooge came up in the xls sheet in 10,000= 1.84 % frequency which is the frequency of 0000 being used as the pin in a set of 10000 pins chosen (according to the website) Pin that belongs to Ebenezer Scrooge is 0000.

Problem 4: The NYC Taxi Authority recently released records of taxi trips in which medallion numbers were concealed via hashing with MD5. (See the Ars Technica article at http://bit.ly/1o0196n.) You can find a full list of cur- rent medallion numbers at https://data.cityofnewyork.us/Transportation/Current- Medallions/avwq-z233. (So can a potential attacker. . .)
Question 4.1: What is the maximum number of attempts needed to crack the hash of a medallion number (assuming no precomputation and no database of issued medallions)? (Hint: The number of possible medallion numbers reported in the Ars Technica article is wrong. Another hint: There are three authorized medallion formats, DLDD, LLDDD, and LLLDDD, where D stands for 'digit' and L for 'letter'.) [5 points]

26x10x10x10+ 26X26X26X1000+26X26X1000 = 18278000

Question 4.2: What is the security goal of the NYC Department of Transporta- tion in this setting, i.e., why did they hash medallion numbers? [3 points]
Confidentiality : even if hashed medallion numbers were leaked, it would be difficult to get back the original medallions numbers from the hashed ones.

Question 4.3: On the course site (file medallionhashes.txt), you'll find a list of SHA-256 hashes for some LLLDDD-format medallions that are active at the time of this exercise. Crack them! Also indicate who owns the corresponding taxis. (Hint: To verify the correctness of the SHA-256 code you use, here's the hash of medallion SBV381:
Code

```
import hashlib
```

```
import csv

hash_file = open('medallionhashes.txt','r')
medallion_file = open('Relevant_Medallion.csv', 'rb')
hash_list = hash_file.readlines()
medallion_list = csv.reader(medallion_file)

for medallion_info in medallion_list:
    for medallion_hash in hash_list:
        if hashlib.sha256(medallion_info[0]).hexdigest() in medallion_hash:
            print medallion_info, medallion_hash
```

OUTPUT

```
['SBV120', 'SINKERIA INC.', 'Stand By Vehicle', 'Unspecified Dirver', '120SBVB',
'2FAFP70W97X104409', '', '2007', 'M', '113']
8f96d287b6b77ed0effdeaa719998894dcc777accb1dbde741b58d14e56957d6

['SBV130', 'OPO TRANSIT INC.', 'Stand By Vehicle', 'Unspecified Dirver', '130SBVB',
'JTDKN3DU7A1301837', 'HYB', '2010', 'M', '312']
daf7123cf1a0ea71c62e174a6290c23d9cb768fae74bb006340ecdfb7d90becb

['SBV132', 'OPO TRANSIT INC.', 'Stand By Vehicle', 'Unspecified Dirver', '132SBVB',
'JTDKN3DU9A1312306', 'HYB', '2010', 'M', '312']
5c2ecc995d856ead993ccdeec1a5163c0bd0d0c1c73929ffef65021b0a5dae0a

['SBV145', 'DHARMA MGT. CORP.', 'Stand By Vehicle', 'Unspecified Dirver',
'145SBVB', '1N4AL2AP6CN431409', '', '2012', 'M', '17']
c89b9b1a6cffd1972ab94ef5dc0e2b3371d98c56ae2c45524e81a2a19fee9be0

['SBV169', 'SBV TAXI CORP', 'Stand By Vehicle', 'Unspecified Dirver', '169SBVC',
'2FAHP70V19X127038', '', '2009', 'M', '11']
ebd0f398d465cc86447c014e9ad4e2060ae4b82314ea84e3787a15d7c2b5ab17

['SBV181', 'JAC SBV CORP', 'Stand By Vehicle', 'Unspecified Dirver', '181SBVB',
'2FABP7AV6AX112995', '', '2010', 'M', '325']
4cd7335fa467de24b767c53e3cfc1789c23e2c36952e66b386fb2ab1b8385066

['SBV192', 'OMFG TRANSIT LLC', 'Stand By Vehicle', 'Unspecified Dirver',
'192SBVB', '4T1BB3EK1BU130892', 'HYB', '2011', 'M', '326']
57f86a9736b1d3ffcfdd15b7a94318ec2ddcab0c5f227a2f7b06cc188feb1287

['SBV265', 'FOREGO TAXI CORP', 'Stand By Vehicle', 'Unspecified Dirver',
'265SBVB', '2FAHP70V59X100067', '', '2009', 'M', '13']
1de578ecf0fd26864f9fcb4e728bcaba839e47d42bbbaaa7b7c62de854110153

['SBV376', '3511 SYSTEMS INC', 'Stand By Vehicle', 'Unspecified Dirver', '376SBVH',
```

'2FABP7AVXAX129623', '', '2010', 'M', '327']
99329a502dd9178b75f3eff01a52555ed1ea9fdbb1a573e47a4adb05f719047a

['SBV379', '3511 SYSTEMS INC.', 'Stand By Vehicle', 'Unspecified Dirver', '379SBVB', '2FABP7AVXBX101662', '', '2011', 'M', '327']
618ecd0a76d5658991e14bc6ef0bbced6ade085b152a32853786dd68156de906

QUESTION 4.4: SUPPOSE THAT VIN NUMBERS AND LICENSEE NAMES WERE CONCATE- NATED WITH MEDALLION NUMBERS PRIOR TO HASHING ACTIVE MEDALLION NUMBERS. WOULD THIS FIX THE PROBLEM AND ACHIEVE THE DESIRED SECURITY GOAL? WHY OR WHY NOT? [5 POINTS]
No it would still not reach the security goal, as both the VIN and the licensee names are also present in an open db, and you can do a concatenation of two, hash them and keep them in a dictionary with the actual values and hashed values and then do a match with the hashed values that have been obtained if the system is hacked.

QUESTION 4.5: DESIGN A CRYPTOGRAPHIC SCHEME (A MECHANISM) USING SHA-256 THAT WOULD PERMIT NYC TO GENERATE A UNIQUE PSEUDONYM FOR EACH TAXI MEDALLION AND MEET THE DESIRED SECURITY GOAL. (NOTE: GIVEN NYC'S DIFFICULTIES, MAYBE THERE'S A STARTUP OR AT LEAST AN INTERESTING PROJECT SOMEWHERE IN HERE. . .) [5 POINTS]
Add a randomly generated salt for each user, hash combination of the salt key and the medallion number and then store it.
So even if the hash of the combination of the salt key and medallion numbers is leaked/hacked, a simple dictionary brute force checking won't work to figure it out.

QUESTION 5.1: OTIS'S FRIEND, MISSISSIPPI MABEL, THE FAMOUS SHARPSHOOTER, SHOT DOWN HIS IDEA. SHE SAID HIS PROTOCOL WON'T WORK BECAUSE THE SERVER CAN CHEAT WITHOUT BEING DETECTED. WHAT IS THE SECURITY GOAL OF THE CLIENT? [2 POINTS] HOW CAN THE SERVER DEFEAT THIS SECURITY GOAL? [2 POINTS] IS THERE ANY WAY IN OTIS'S PROTOCOL FOR THE CLIENT TO DETECT SERVER CHEATING? [2 POINTS]
Security goal is integrity.
The server can cheat the client by taking the cslot given by client and keep running the random generator, computing the modulus till it gives a different value than the value entered by the client, and hence never allow the client to win. No there is no way in this protocol for the client to know the server is cheating.

QUESTION 5.2: MABEL CAME UP WITH A CLEVER PROTOCOL TO ENABLE THE CLIENT TO DETECT A CHEATING SERVER IMMEDIATELY. WHY DOESN'T MABEL'S PROTOCOL WORK, I.E., HOW CAN A SERVER ENSURE THAT IT NEVER AWARDS A WIN TO A CLIENT BUT CHEATED = FALSE? [5 POINTS]
Mabel's Protocol does not work because the key is known to the server, hence it can just hash the combination of the key as well as the different values of the slot ie 0 to 38

and create a dictionary, and then then compute the cslot (slot number of the client) and then deliver a different sslot, hence makes it always such that the client never wins.

QUESTION 5.3 (BONUS): FIX MABEL'S PROTOCOL SO THAT IT DOES THAT DOES WORK, I.E., IT ALLOWS A CLIENT TO DETECT UNFAIR PLAY IMMEDIATELY. [10 POINTS]

Mabel's Protocol can be fixed by using a salt key which is unique for each transaction, hashing it along with the key and the cslot, and then allowing server to calculate the slot, and then telling the server the cslot, along with the salt key.

C: Salt->Random number

C: a = SHA256(cslot // K || Salt)

C→S: a

S: y = truerand();

sslot = y mod 39

S→C: sslot

C→S: cslot

C->S: Salt

S: if a = SHA256(cslot // K || Salt) AND cslot = sslot then

win = true

else

win=false

S→C: win

C: if cslot = sslot AND win = false

cheated=true

else

cheated = false

PROBLEM 6 (BONUS): RECALL THAT A MAC IS A "TAG" THAT MAY BE APPENDED TO A MESSAGE (PLAINTEXT OR CIPHERTEXT) TO ENSURE THAT TAMPERING CAN BE DETECTED BY
4
S→C: WIN C:
WIN = FALSE
IF CSLOT = SSLOT AND WIN = FALSE
THE RECEIVER. MACS ARE TYPICALLY CONSTRUCTED USING HASH FUNCTIONS. A COMMON EXAMPLE IS HMAC.
THE CHIEF CRYPTOGRAPHER OF THE RNB USED SHA-256 TO DESIGN A NEW, CUS- TOM MAC CALLED RURIMAC. UNFORTUNATELY, WHILE RURIMAC CANNOT FORGED BY AN EAVESDROPPER THAT DOESN'T INTERCEPT A VALID (MESSAGE, MAC) PAIR, AND DOES MAKE RECOVERY OF THE SECRET KEY INFEASIBLE, IT CAN BE FORGED BY AN ADVERSARY THAT INTER- CEPTS A SINGLE VALID (MESSAGE, MAC) PAIR. GIVE A POSSIBLE DESIGN FOR RURIMAC. [10 POINTS]

Using Length extension attacks, if certain types of hashes are used as message authentication codes, allow inclusion of extra information. This attack can be done on SHA-256 hashes with construction H(secret ∥ message) when message and the length of secret is known.
Since HMAC doesn't use the construction H(key ∥ message), HMAC hashes H(key ∥ H(key ∥ message)) even using susceptible algorithms are not prone to length extension attacks