

Name: Gan Chong Yee
Title: Homework 1 – Security and Privacy Concepts in the Wild

Problem 1

1. That high-profile Twitter account might be linked to the victim's gmail and other useful accounts, such as Amazon. By pivoting through different accounts, the hacker would be able to access the victim's credit card information to obtain money or other benefits worth more than \$325.
2. The hacker might be interested in defaming the victim due to their high-profile status.
3. The hacker can use social engineering to send phishing links and malware to close contacts of the victim. This allows the hacker to obtain even more passwords from the victim's social network for other benefits and monetary gain.

Problem 2

Question 2.1 Statistics from census.gov in the year 2000 was used to compute the guessing probability (GP). As the guessing probability password (GPP) is "Smith"¹, we assume the hacker would use the GPP to attack multiple accounts sequentially – they have the highest incentive to do so as it gives the best probability of success, assuming there is no security measures against the number of accounts they can try this on. With the total population in year 2000 shown in Figure 1, the formula is thus:

$$\begin{aligned} GP &= \frac{\text{frequency}_{\text{smith}}}{\text{total population}} \\ &= \frac{2376206}{281421906} \\ &= 0.0084 \end{aligned}$$

$$\begin{aligned} \text{Number of tries} &= \frac{1}{GP} \\ &= 119 \end{aligned}$$

¹ http://www.census.gov/topics/population/genealogy/data/2000_surnames.html

Table 1. Total Population by Age, Race and Hispanic or Latino Origin for the United States: 2000

Internet Release Date: October 3, 2001

[For information on confidentiality protection, nonsampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf1.pdf>.]

		Race								
		One race								
					American Indian and Alaska Native		Native Hawaiian and Other Pacific Islander			
	Total population	Total	White	Black or African American		Asian		Some other race	Two or more races	Hispanic or Latino (of any race)
FIVE-YEAR AGE GROUPS										
Total population	281 421 906	274 595 678	211 460 626	34 658 190	2 475 956	10 242 998	398 835	15 359 073	6 826 228	35 305 818
Under 5 years	19 175 798	18 227 583	12 859 892	2 804 786	213 052	670 406	33 391	1 646 056	948 215	3 717 974
5 to 9 years	20 549 505	19 719 732	13 944 882	3 205 512	239 007	680 536	36 503	1 613 292	829 773	3 623 680
10 to 14 years	20 528 072	19 824 608	14 322 638	3 121 530	245 677	684 525	35 772	1 414 466	703 464	3 163 412
15 to 19 years	20 219 890	19 597 998	14 167 148	2 929 553	232 351	746 511	37 328	1 485 107	621 892	3 171 646
20 to 24 years	18 964 001	18 411 917	13 064 891	2 628 752	198 010	816 452	38 693	1 665 119	552 084	3 409 427
25 to 29 years	19 381 336	18 868 887	13 501 773	2 548 968	186 689	986 222	35 224	1 610 011	512 449	3 385 334
30 to 34 years	20 510 388	20 026 210	14 818 786	2 618 602	186 072	949 418	33 129	1 420 203	484 178	3 124 901
35 to 39 years	22 706 664	22 235 945	17 031 493	2 826 361	202 013	909 439	33 031	1 233 608	470 719	2 825 158
40 to 44 years	22 441 863	22 021 176	17 265 995	2 700 418	189 201	846 118	28 760	990 684	420 687	2 304 152
45 to 49 years	20 092 404	19 754 156	15 810 626	2 275 191	159 422	749 777	23 675	735 465	338 248	1 775 168
50 to 54 years	17 585 548	17 316 932	14 213 875	1 805 457	128 303	626 255	18 938	524 104	268 616	1 360 935
55 to 59 years	13 469 237	13 280 566	11 107 247	1 306 641	90 531	433 749	13 428	328 970	188 671	960 033
60 to 64 years	10 805 447	10 662 421	8 945 842	1 063 469	67 189	342 795	10 142	232 984	143 026	750 407
65 to 69 years	9 533 545	9 421 591	8 040 225	881 786	49 463	274 085	7 698	168 334	111 954	599 353
70 to 74 years	8 857 441	8 766 843	7 648 193	731 386	36 434	220 066	5 529	125 235	90 598	477 266
75 to 79 years	7 415 813	7 348 823	6 530 019	550 024	25 608	155 965	3 614	83 593	66 990	326 726
80 to 84 years	4 945 367	4 904 714	4 408 597	346 465	14 646	88 183	2 155	44 668	40 653	179 538
85 years and over	4 239 587	4 205 576	3 778 504	313 289	12 288	62 496	1 825	37 174	34 011	150 708
SELECTED AGE GROUPS										
Under 18 years	72 293 812	69 436 926	49 598 289	10 885 696	840 312	2 464 999	127 179	5 520 451	2 856 886	12 342 259
Under 1 year	3 805 648	3 602 103	2 535 928	548 955	42 167	129 803	6 464	338 786	203 545	771 053
1 to 4 years	15 370 150	14 625 480	10 323 964	2 255 831	170 885	540 603	26 927	1 307 270	744 670	2 946 921
5 to 13 years	37 025 346	35 623 089	25 411 015	5 727 934	436 694	1 227 263	65 181	2 755 002	1 402 257	6 185 947
14 to 17 years	16 092 668	15 586 254	11 327 382	2 352 976	190 566	567 330	28 607	1 119 393	506 414	2 438 338

Figure 1: Statistics of Total Population for year 2000

Question 2.2 The security question “In what city did your parents meet” was found in the sign up page for an AppleID, as shown in Figure 2. From City Mayors Statistics², it can be seen that the city with highest population in 2010 was New York City. With the total population of year 2010³:

$$\begin{aligned}
 GP &= \frac{\text{population}_{NYC}}{\text{total population}} \\
 &= \frac{8175133}{308745538} \\
 &= 0.0265
 \end{aligned}$$

$$\begin{aligned}
 \text{Number of tries} &= \frac{1}{GP} \\
 &= 38
 \end{aligned}$$

² http://www.citymayors.com/gratis/uscities_100.html

³ <http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>

Apple ID and Password

Enter your primary email address as your Apple ID. This will be used as the contact email address for your account.

Apple ID	<input type="text" value="example: jappleseed@example.com"/>
Password	<input type="password"/>
Confirm Password	<input type="password"/>

Security Questions

Select three security questions below. These questions will help us verify your identity should you forget your password.


Security Question	<div>Please select </div>
Answer	<input type="text"/>
Security Question	<div>✓ Please select What was the first name of your first boss? In what city did your parents meet? What was the name of your first pet? What is the first name of your best friend in high school? What was the first film you saw in the theater? What was the first thing you learned to cook?</div>
Answer	<input type="text"/>

Figure 2: Security Question for AppleID – In what city did your parents meet?

Problem 3

Question 3.1 Assuming that there is no security measures against trying out the Pin unlimited number of times, the attacker can find out the pin with a probability of 1 if he try all possible pin combinations. This is simply $10 \times 10 \times 10 \times 10 = 10,000 = 10^4$ attempts.

Question 3.2 The number of encrypted data that matches Ebenezer Scrooge's was calculated in the "RNB PINs.xlsx" sheet using the formula:

`=COUNTIF(B5:B10004,"="&B921)`

The frequency is 184, as seen in Figure 3. This accounts for $184/10000 \times 100 = 18.4\%$ of the list of encrypted data. As the encryption was performed on 4-digit pins, this also means Scrooge's PIN is approximately the 18.4% most used 4-digit pin. As can be seen from Figure 4, obtained from the blogpost⁴, this shows that Ebenezer Scrooge's PIN code is probably '0000', as it is the closest in percentage as can be seen in Figure 4.

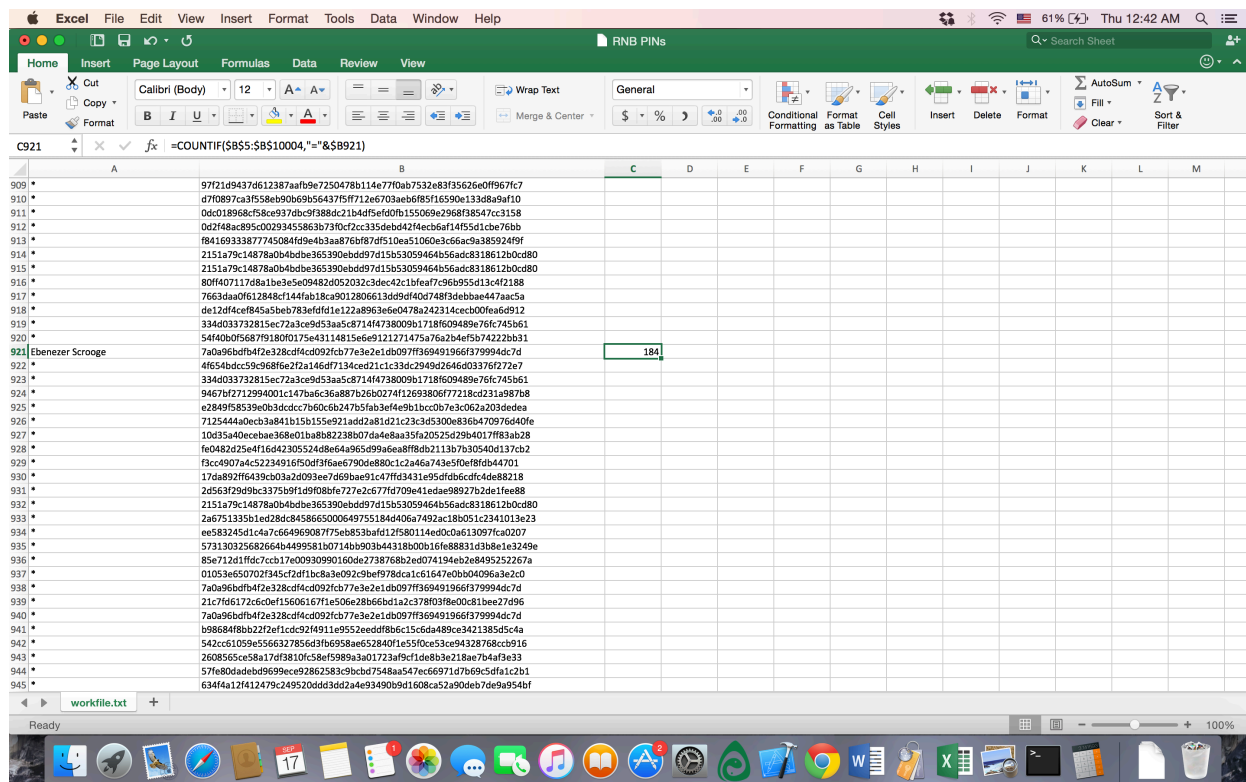


Figure 3: Frequency of Scrooge's encrypted PIN calculated from the sheet

⁴ <http://www.datagenetics.com/blog/september32012/>

	PIN	Freq
#1	1234	10.713%
#2	1111	6.016%
#3	0000	1.881%
#4	1212	1.197%
#5	7777	0.745%
#6	1004	0.616%
#7	2000	0.613%
#8	4444	0.526%
#9	2222	0.516%
#10	6969	0.512%
#11	9999	0.451%
#12	3333	0.419%
#13	5555	0.395%
#14	6666	0.391%
#15	1122	0.366%
#16	1313	0.304%
#17	8888	0.303%
#18	4321	0.293%
#19	2001	0.290%
#20	1010	0.285%

Figure 4: Table of most frequently used 4-digit PINs

Problem 4

Question 4.1 The maximum number of attempts is simply the total number of possible medallion numbers. As the authorized medallion formats are DLDD, LLDDD and LLLDDD, with L being one of 26 possible alphabets, and D being one of 10 possible digits, the numberthe total number of ways = $26 \cdot 1000 + 26 \cdot 26 \cdot 1000 + 26 \cdot 26 \cdot 26 \cdot 1000 = 18,278,000$ attempts.

Question 4.2 To ensure the privacy of taxi drivers. Should the medallion number be known, hackers can find out the name, income, work performance and previous trips made by the driver. Worse still, their security can be breached, as it would be easy to discover their address and future locations from these data.

Question 4.3 The code for the cracking the hashes was attached with the assignment as hw1_ques4.4_securities.ipynb. The medallion hashes, medallion numbers and names of licensee respectively were shown in Figure 5:

```
[ '8f96d287b6b77ed0effdeaa719998894dcc777accb1dbde741b58d14e56957d6', 'SBV120', 'SINKERIA INC.' ]
[ 'daf7123cf1a0ea71c62e174a6290c23d9cb768fae74bb006340ecd7b7d90becb', 'SBV130', 'OPO TRANSIT INC.' ]
[ '5c2ecc995d856ead993ccdeec1a5163c0bd0d0c1c73929ffef65021b0a5dae0a', 'SBV132', 'OPO TRANSIT INC.' ]
[ 'c89b9b1a6cfff1972ab94ef5dc0e2b3371d98c56ae2c45524e81a2a19fee9be0', 'SBV145', 'DHARMA MGT. CORP.' ]
[ 'ebd0f398d465cc86447c014e9ad4e2060ae4b82314ea84e3787a15d7c2b5ab17', 'SBV169', 'SBV TAXI CORP' ]
[ '4cd7335fa467de24b767c53e3cfc1789c23e2c36952e66b386fb2ab1b8385066', 'SBV181', 'JAC SBV CORP' ]
[ '57f86a9736b1d3ffcfdd15b7a94318ec2ddcab0c5f227a2f7b06cc188feb1287', 'SBV192', 'OMFG TRANSIT LLC' ]
[ '1de578ecf0fd26864f9fcb4e728bcaba839e47d42bbbaaa7b7c62de854110153', 'SBV265', 'FOREGO TAXI CORP' ]
[ '99329a502dd9178b75f3eff01a52555ed1ea9fdbb1a573e47a4adb05f719047a', 'SBV376', '3511 SYSTEMS INC' ]
[ '618ecd0a76d5658991e14bc6ef0bbced6ade085b152a32853786dd68156de906', 'SBV379', '3511 SYSTEMS INC.' ]
```

Figure 5: Results of Cranking – medallion hash, medallion number, Licensee name

Question 4.4 No, if the hacker had access to the released data⁵. If he knew beforehand that the preimage consists of a concatenated version of VIN numbers, licensee names and medallion number, the code can be cracked by simply concatenating the corresponding database elements before running the program. However, if the hacker is unaware of the way these elements are concatenated, or if they are concatenated by random, the total number of possible preimage would increase, making it slightly harder to crack:

Total number of preimage = number of instance * 3P3
= number of instance*6

However, if the hacker had no access to the data³, it would be a lot harder but still crackable. The total number of possible combinations will raise by 10^{17} for possible VINs (there are 17 digits), and by many more orders of difficulty from the number of possible names (the names are of different lengths of string, and some of them are human names whilst the rest are corporation names). This can be easily overcome if the attacker finds some way to obtain the full list of taxi owners. In both cases, though, it is still possible to crack the code given enough time and processing power.

Question 4.5 NYC can utilize a unique key to append to the medallion number prior to hashing it. This would make it almost impossible for the attacker to code, even with a 256-bit randomly generated key. In the same way, the names and VINs should also be encrypted with a key to prevent the attacker from obtaining or inferring useful information for hacking. The input can also be hashed multiple times, or a salted password can also be used to prevent cracking.

Problem 5

Question 5.1 The security goal of the client is to ensure that the server does not cheat and change slot after it receives slot. This can be done if the function `y – truerand()` is not truly random, and is a specific number that does not match slot instead. The client cannot ever detect the server cheating, unless he has access to the server's codes.

⁵ <https://data.cityofnewyork.us/Transportation/Current-Medallions/avwq-z233> or <https://bit.ly/1ITn06z>.

Question 5.2 The server can ensure that the client will never win but still have cheated = false by simply using brute force to guess what cslot is from a. Since the server knows the key as well, he can easily acquire all 39 possible outputs of the function by passing 39 possible values of cslot (there are 39 integer values) into $\text{SHA256}(\text{cslot} || K)$. From there, he can check which output matches a, and deduce cslot from it. He can thus assign sslot with a value that is not equal to cslot, and pass it to the client unnoticed.

Question 5.3 The client can hide the key, K, from the server first when sending a. The client will then send K to the server only after he has received sslot from the server. This way, both party can check if they have cheated. This is because if the client cheats by changing the K later on, the server can simply check all 39 possible outputs of that K. If the possible outputs doesn't match the a previously passed by the client, it means the client is cheating.