Jennifer Shi
j77shi 20662230 j77shi@uwaterloo.ca CS458

a3-responses

Question 1: Diffie-Hellman Key Exchange [9 marks]

(a) [3 marks] Assume Alice and Bob have already agreed to use modulus *p* = 107 and base *g* = 17.
Then Alice chooses secret parameter *a* = 12 and Bob chooses secret parameter *b* = 34. What is the public value *A* that Alice gives to Bob? What is the public value *B* that Bob gives to Alice? What is the resulting secret key that is generated as a result of DH protocol?

Public value A that Alice gives to Bob:

$A = g^{a} \text{ (mod p)} = 17^{12} \text{ (mod 107)} = 34$

Public value B that Bob gives to Alice:

$B = g^{b} \text{ (mod p)} = 17^{34} \text{ (mod 107)} = 29$

Secret key that is generated as a result of DH protocol:

$K = g^{ab}\text{(mod p)} = 17^{34*12}\text{(mod 107)} = 47$

(b) [2 marks] During the key exchange, Eve observes *g*, *p*, $A = g_a$ (mod *p*), and $B = g_b$ (mod *p*). Can Eve recover the original secret values *a* or *b* using this information? Explain why or why not.

For this question specifically, Eve will be able to recover the original secret values a or b because p is too small, so it is easy to guess a and b by using the discrete logarithm.

In general, Eve cannot recover the original secret values a or b using this information, because when p is properly chosen, it is computationally infeasible to solve for a and b. When p is large enough, Eve wouldn't be able to recover the original secret values of a or b.

(c) [2 marks] What if Mallory comes along and behaves as an active Man-In-The-Middle (MITM) attacker, how can she manipulate the Diffie-Hellman protocol to obtain all of the plaintext communications between Alice and Bob? Explain.

Mallory can act as the MITM by establishing a key with Alice and Bob respectively.

Alice sends her key $A=g^{a}$ mod p to Bob, and Mallory intercepts it and replaces it with g^{m} mod p.

Bob sends his key $B=g^{b}$ mod p to Alice, and Mallory intercepts it and replaces it with g^{m} mod p.

Mallory can compute the shared keys using the above information from Alice and Bob. $(g^a \bmod p)^b \bmod p$, and $(g^b \bmod p)^a \bmod p$ respectively.

So, when Alice sends Bob message (or Bob sends Alice message), Mallory will be able to decrypt and alter the message, and neither of them will know.

(d) [2 marks] Provide a brief explanation of how Alice and Bob could prevent such an attack by Mallory.

A way to prevent this attack by Mallory is to add an additional layer of security, for example, let g^a and g^b be signed by a Certificate Authority or adding digital signature so that it can be made sure that is not being altered when receiving it.

Question 2: Textbook RSA [7 marks]

In RSA, the public key is a pair of integers (*n; e*), where *n* = *pq* for large primes *p* and *q*. The private key is the triple (*p; q; d*) where *de* ≡ 1 mod (*p* - 1)(*q* - 1). In a simplified form of RSA, called "textbook RSA", the encryption of a message *m* to yield the ciphertext *c* is $c = m_e$ mod *n* and the decryption is $m \equiv c_d$ mod *n*.

(a) [2 marks] Does textbook RSA provide semantic security? If so explain why, if not provide a simple countermeasure.

RSA does not provide semantic security, because it is deterministic. For example, we can look at the encryption for 0 and 1 and found out that there are patterns to the ciphertexts.

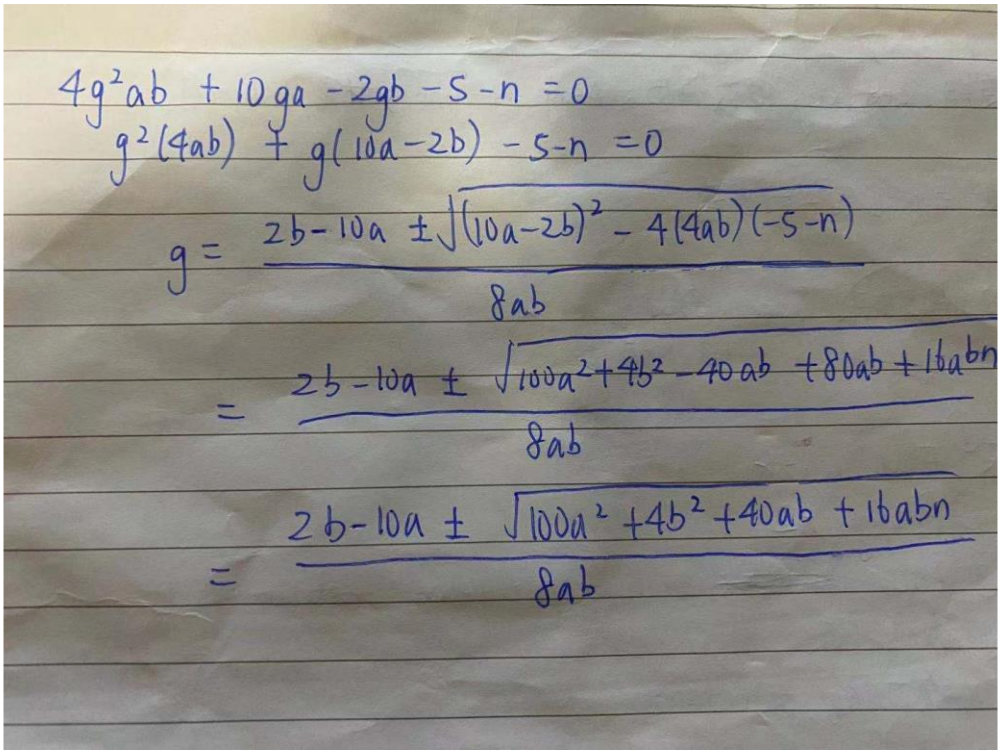
(b) [2 marks] Write the number *n* in terms of *g*, *a* and *b*.

p = 2ga – 1

q = 2gb + 5

n = pq = (2ga-1)*(2gb+5) = $4g^2$ab + 10ga-2gb – 5

(c) [2 marks] Use the above relation to give a closed-form expression for *g*.



$$4g^2ab + 10ga - 2gb - 5 - n = 0$$

$$g^2(4ab) + g(10a-2b) - 5 - n = 0$$

$$g = \frac{2b - 10a \pm \sqrt{(10a-2b)^2 - 4(4ab)(-5-n)}}{8ab}$$

$$= \frac{2b - 10a \pm \sqrt{100a^2 + 4b^2 - 40ab + 80ab + 16abn}}{8ab}$$

$$= \frac{2b - 10a \pm \sqrt{100a^2 + 4b^2 + 40ab + 16abn}}{8ab}$$

(d) [1 marks] Write in one sentence, how the above information helps you in finding the factors of n.

Since we are given a and b. And we know that $n=pq$, $p = 2ga -1$, $q = 2gb + 5$. We can simply plug in values for a and b in the relation we derived in part c and solve for p and q.

Question 3: Tracker Attacks [7 marks]

(a)

i) Your assumption.

Assume that the number of part time and full-time workers are between k and N-k (or N/8 and 7N/8). This assumption is realistic because it is reasonable that in a local police department to have both full time and part time workers, and the ratio 1:8 or above is reasonable.

ii) Your Tracker.

Let the tracker q(T) be the query SELECT SUM(Salary) FROM Employee WHERE Type = "Full Time"

iii) The set of 3 queries you will use.

1. q (C or T) = SELECT Sum(Salary) FROM Employee WHERE Type = "Full Time" or Name = "Olivia"
2. q (C or not T) = SELECT Sum(Salary) FROM Employee WHERE Type <> "Full Time" or Name = "Olivia"
3. q(S) = SELECT Sum(Salary) FROM Employee

iv) How you can use the results of the three queries to obtain Olivia's Salary.

Olivia's salary: Q(C or T) + q(C or not T) – q(S)

(b)

Let Y be the salary that we are guessing Sarah will have. Take the range of Y to be 0 to 200,000.

1. q(C or T) = SELECT COUNT(*) FROM Employee WHERE Name= "Sarah" AND Salary = Y OR Salary >Avg(Salary)
2. q(C or not T) = SELECT COUNT(*) FROM Employee WHERE Name= "Sarah" AND Salary = Y OR Salary <=Avg(Salary)
3. q(S) = SELECT COUNT (*) FROM Employee WHERE Salary > 0

If q(C or T) + q(C or not T) – q(S) = 1, it means that Sarah has salary Y and our guess was correct.

Question 4: Naive Anonymization [10 marks]

a) Cathy:7; We know that Cathy is 7 because she emails everyone except 4.

Ed: 4; because 4 is the only one that Cathy (7) is not emailing

Alice: 6; because Ed only emails 5 and 6, and 6 emails three people(4,5,7), while 5 emails 4 people (3,4,6,7)

Bob: 5; deduction from Alice's list, since Alice emails 4,5,7 and 4 is Ed and 7 is Cathy, so Bob must be 5

b) I think they will be able to break this anonymization because they are using a fixed secret linear perturbation function (ax+b=y) to add noises to the data. When Alice and Cathy team up, they will be able to generate two equations, we can solve the two unknowns a and b.

Specifically, from Alice :14 = 32a + b

From Cathy: 64 = 57a + b

Solving the system of equations, we will get a=2 and b = -50.

We will be able to obtain the true number of emails by substituting a and b and the number of emails in the table to obtain the true number of emails.

c) False. A 3-anonymized table must have at least 2 other users with indifferent traits. The table is not 3-anonymous because there are only two records for Net Income [15-28].

In order to be 3-anonymized, we can generalize Net Income to: [-13 - -1], [0-12], [13-25].

There are at least 3 "well-represented" values of the sensitive information (Manufacturing Facility) for each quasi-identifier, thus the table is 3-diverse.

d) We will know that Cactus Costumes' Manufacturing Facility is Boston since it is -1 and it is on both tables. We know from the second table the range [-14-0] that Cactus Costume's Manufacturing Facility is in one of Boston, Toronto and Vancouver. And we know that only Boston is listed in Table 1. So, we can deduce that Cactus Costumes' Manufacturing Facility is in Boston.

e) We can randomly perturbate some data, for example, swap the value of some data, thus trade off some accuracy with privacy to resist attackers with background knowledge.