Connor Raymond Stewart
20673233

CS658 Assignment Three

Question 1:

1. In a TLS handshake, the client and server agree on information such as the cipher suites to use and session keys for symmetric encryption. The client authenticates the server's identity though the use of SSL certificates digital signatures and the servers public key, and the client sends a premaster secret which it encrypts using the servers public key so that the server can decrypt the premaster secret with its private key.
2. Asymmetric encryption establishes a secure connection between the server and the client. Symmetric encryption allows exchanging data during a secure connection established through an asymmetric encryption-based session.
   Asymmetric key encryption uses a public key for encryption and a private key for decryption, whereas symmetric key encryption uses only a single key for both encryption and decryption. Asymmetric key encryption allows public keys to be shared without needing to worry about security of the keys, however the mathematically property of asymmetric key encryption means that keys need to be much larger than symmetric keys, which makes increases computational complexity for encryption operations. Symmetric key cryptography has only one key, meaning that the key cannot be transferred without having some type of security, but it is very fast compared to asymmetric key cryptography. Therefore, asymmetric encryption can be used to establish a connection to deliver a symmetric key which is used for communication afterwards.
3. Certificate authorities are entities that confer digital certificates. A digital certificate certifies ownership of a public key by the subject of the certificate. Certifying the ownership of a public key from a server can be used as a mechanism to support the authenticating of servers.
4. Man in the middle attack:
   a. Firstly, Mallory needs to intercept Alice's request for Bobs public key which Alice requests when she attempts to connect to Bobs website. When Mallory captures Alice's public key request to Bobs website, he inserts his address into the response address and sends the request to the server. The server will reply to Mallory with a copy of its public key to Mallory, and Mallory sends his public key to Alice to make himself an intermediate man in the middle. Mallory then generates a premaster secret with Bobs public key, and Alice generates a premaster secret with Mallory's public key. Alice's browser sends its premaster secret it made using Mallory's public key to Mallory's computer; and Mallory's computer then sends its premaster secret to the server that Alice's browser thinks its connecting to. The server and Mallory, along with Alice and Mallory; generate a master secret for the connection they created, meaning Alice's browser thinks it has a secure connection to Bob's server, but Bob's server and Alice's browser are both connected to Mallory. Therefore, Mallory can capture and read everything Alice's browser and Bob's server transmit.
   b. In addition to the general steps listed above, a specific set of steps include:

i. Mallory needs to intercept and capture Alice's request for Bob's public key and replaces it with his own request.
ii. Mallory needs to intercept and capture Bobs public key to Mallory and replace it with his own public key.
iii. The client uses Mallory's public key to generate a premaster secret which is sent to Mallory, who in turn generates a premaster secret to Bob using the server's public key.
iv. The server and Mallory generate a master secret, and Mallory and Alice generate a master secret. Alice now thinks she is connected to the server, but is communicating with Bob, and Bob is communicating with the server on her behalf. The master secrets will be used for symmetric key encryption in the following steps.
v. Alice attempts to send her banking credentials Bob's server and encrypts it with the Mallory-Alice master secret pair, but instead sends it to Mallory via the man in the middle attack started above. Mallory decrypts Alice's information using his master secret pairing with Alice generated in step iv.
vi. After intercepting the banking credentials from Alice and decrypting them using the Alice-Mallory master secret pairing, Mallory saves the credentials and therefore *steals the banking information*. However, Mallory is not finished since he needs a purchase to go through to convince Alice nothing is wrong. Mallory encrypts the banking credentials using his Mallory-Bob master secret pairing and sends the credentials to Bob's server.
vii. Bob's server sends purchase confirmation information back to Mallory and encrypts it using the Mallory-Bob master secret pairing, and Mallory decrypts it using his copy of the master secret pairing. Mallory then encrypts the data using his copy of the Alice-Mallory master secret pairing, and Alice decrypts it using her copy of the Alice-Mallory master secret pairing. The man in the middle attack connection can now be terminated as Alice logs off the website.

Question 2:

a) We can conclude the identities:
   a. We note that Alice has four connections, so she should be able to connect to either note 4 or 5. However, we also know that Bob and Dave have the same number of contacts as each other and more then Cathy. Since Node 5 is connected to only two nodes (3 and 8) with equal values of one contact, node 5 cannot be Alice since Bob and Dave would not be able to have more contacts than Cathy. Therefore, Alice is located at node 4.
   b. Node 4 has two nodes (1 and 6) with equal contact values of three. We note that Cathy must have less contacts than Dave and Bob, meaning node 2 with two contacts is the only suitable candidate connected to Alice on node 4. Therefore,

Cathy is located at node 2, Robert must be located at node 5 since it is the last possible node to connect to, and Bob and Dave are located on nodes 1 and 6.

c. Finally, we know that Cathy contacts with neither Dave nor Robert, meaning Cathy must connect to bob at node 1. Therefore, we can isolate Bob to node 1 and Dave to node 6. We see that Alice must connect – and actively does connect to – Bob, Cathy, Dave, and Robert.

d. In conclusion: {Alice = 4, Robert = 5, Dave = 6, Bob = 1, Cathy = 2}

b) We can use algebra to determine the values of m and b. By setting both equations to equal b, we can use algebraic manipulation to make both formulas set to each other. Since both formulas are modeled by the same equation, they can used to determine the value of m and then m's value can be used to calculate b directly:

Formulations
Equation one
23 = 24m + b
b = 23 – 24m
Equation two
31 = 56m + b
b = 31 – 56m
Eq1 = Eq2
23 – 24m = 31 – 56m
-24m + 56m = 31 – 23
32m = 8
m = 8 / 32 = 1 / 4 = 0.25
Isolate for b
23 = 24(0.25) + b
b = 23 – 24(0.25) = 23 – 6 = 17
AND:
31 = 56m + b
b = 31 – 56m = 31 – 56(0.25) = 31 – 14 = 17
Therefore, we can conclude that m = 0.25 and b = 17

Now, we note that m = 0.25, b = 17, x = Original Number of Emails (Unknown), and y = Number of Emails Published (Known). We can derive the following equation to obtain x:

y = mx + b → mx = y – b, so:

$$x = \frac{y - b}{m}$$

We calculate the following IDs from the Published table:

ID	Number of Emails	Original Number of Emails

9	31	56
10	36	76
11	24	28
12	25	32
13	37	80
14	27	40
15	23	24
16	21	16

c) The table is 4-annonymous. 4-annonomy is achieved by taking the quasi-identifier attributes of Age, Zip code, and Birth country and pairing them such that every tuple in the table shares its identifier value with at minimum k minus one other values in the table.

d) The table is 3-diverse since three groups based on Age, Zip, and Birth can form a set sharing a combination of attributes such that only the diagnosis may differ between the rows within the sets.

e) Knowing that Matt is a roughly thirty-five-year-old Canadian with a postal code starting with L6A who checked into the hospital on January the third and the eleventh allows us to isolate for *heart disease*. Out of the eight different rows between the two tables for the separate days which satisfy the postal code, citizenship, and age identifiers, only heart disease is present as a consistent diagnosis. Heart disease cannot go away within a short period of time, and therefore Matt must have heart disease.

f) We know there are eight people from the table and that any of them can have up to a maximum of 400 emails per day. We also know that the difference between one value and another is an individual email for a single person. The average is calculated as the sum of all emails across all people over the eight days. Therefore, by taking the maximum and subtracting by one less email gives the sensitivity.

$f(D_1) = (400 * 8) / 8 = 400$

$f(D_2) = (400 * 7 + 399) / 8 = 399.875$

$\Delta f = \max\|f(D_1) - f(D_2)\|_1 = 0.125$

Thus, the sensitivity is 0.125 since changing the emails count by one changes f by 0.125

Question 3:

a) The most trivial solution involves the user requesting the server to send all the information stored in A for download. The q vector consists of n-separate vectors with c ranging from one to n such that all c elements are returned. The server is unable to know what we specifically searched for, since we simply extract everything. Since n vectors need to be uploaded to the server to obtain each row, and since all n vectors contain n-bytes, n^2 bytes need to be uploaded to the server.

b) Since Frieda wants to obtain the details of n/2 products and since each product has m bytes of data, we withdraw (nm)/2 matrix elements.
since n/2 customers get a single product and each product has m bytes of data, we withdraw (nm)/2 matrix elements.

since redundant zeros are passed along with Friedas requests, the trivial solution would be better for downloading. With the customers requesting a single element from the matrix, they are not repeating similar requests multiple times so the trivial solution would increase and duplicate data across all n/2 customers. Frieda would not need to download blank information if she were to simply use the trivial solution.

c) Each q-vector has a byte set to one for the row to be extracted and therefore the s-vectors contain the rows of elements pertaining to the bytes set in the q-vectors. Since the s_1 and s_2 vector both contain elements from a row, they are row vectors of length m. s_3 is a vector consisting of the XOR of the q_1, q_2, and q vectors. If any of the q vectors match, they cancel due to the exclusivity property of the XOR operation. That would mean that q_3 would directly equal one of the q values. In the event one of the q-values is directly set to q_3, then it product a s-vector identical to an s-vector in step six which will cancel out the vectors to let the final s-vector equal one of the s-number vectors. In the alternative case where q_1, q_2, and q are all distinct, q_3 will be the product of all values in the q vectors. In that event, the resultant s_3 vector will contain a sum of three A_{nm} elements from matrix A. We know due to the XOR property the following holds:

s_1 XOR [A_{21}, A_{22}, ….] XOR [A_{11}+A_{21}+A_{c1}, …….]

if A_{c1} = 1 then:

s_1 XOR [$\cancel{A_{21}}$, $\cancel{A_{22}}$, ….] XOR [A_{11}+$\cancel{A_{21}+1}$, …….]

s_1 XOR [A_{11}+1, …….]

[1, ……] where A_{c1} = 1

if A_{c1} = 0 then:

s_1 XOR [$\cancel{A_{21}}$, $\cancel{A_{22}}$, ….] XOR [A_{11}+$\cancel{A_{21}}$, …….]

[$\cancel{A_{11}}$, ….] XOR [$\cancel{A_{11}}$, …….]

[0, ……] where A_{c1} = 0

Thus, in both cases we evaluate the equation to be s. So, in both events for the q values canceling through the XOR operation of not, we get s through the XOR operation in step 6. Therefore, s = q*A = A_c.

d) The servers learn information about the row being viewed, but not the other rows. The servers cannot know if the row being viewed is a decoy row or the real row we are trying to obtain. If the servers communicate with each other, they can algebraically isolate for our true q-vector as can be seen in the proof above. Therefore, the servers cannot yield any useful information without directly communicating with each other.

e) The total upload rate is 3n element, and the download rate is 3m elements.

f) We note an inflection point occurs along n=3 such that the trivial solution requests more matrix elements when n>3. See below:

Trivial upload: n^2; and trivial download: n*m

n = 1:

Trivial upload = 1, trivial download = 15

IT-PIR upload = 3, IR-PIR download = 45

n = 2:

Trivial upload = 4, trivial download = 30

IT-PIR upload = 6, IR-PIR download = 45

n = 3:

Trivial upload = 9, trivial download = 45

IT-PIR upload = 9, IR-PIR download = 45

n = 4:

Trivial upload = 16, trivial download = 60

IT-PIR upload = 12, IR-PIR download = 45

…… (the pattern continues from here)

Thus, the trivial solution is greater for all n > 3 and the trivial solution is equivalent only at n = 3.