Q1:

1. A
2. A

Q2:

(a) This attack will succeed assuming Alice falls for the phishing attack and inputs her login information on the phishing website. This is because Mallory planted a certificate within Alice's browser, so Alice authenticates the server correctly. When Alice logs into the website from the link Mallory sent Alice's browser will not throw any TLS errors and Mallory will have completed her man in the middle attack.

(b) This attack will fail. Mallory is attempting to compromise a CA by planting a fake certificate that she will use to communicate to both the real server and Alice by masquerading as the real server from Alice's perspective. However Mallory mistakenly sent the certificate that was issued by the CA but received by Mallory ($s_k^{CA}, v_k^{MALLORY}$) instead of sending a new forged certificate issued by Mallory (using CA's v_k) with CA's verification like ($s_k^{CA}, v_k^{MALLORY}$).

(c) Mallory's attack will succeed because Mallory's website is the one that is received all the information and Alice's browser has received the certificate that Alice's browser is expecting (the real website's certificate) then Mallory can collect Alices credentials.

(d) A

a. Because TLS uses TCP and the initial handshake phase cannot be completed without using a UDP connection Mallory will need to steal Alice's information (ClientHello) that is being sent to Bob such as the client-key-share/PSK-label and Bob's servers information (ServerHello) such as server-key-share/PSK-label which are both sent in plain text between each other. Mallory can then encrypt, decrypt, sign and verify messages that are being sent between Bob and Alice.

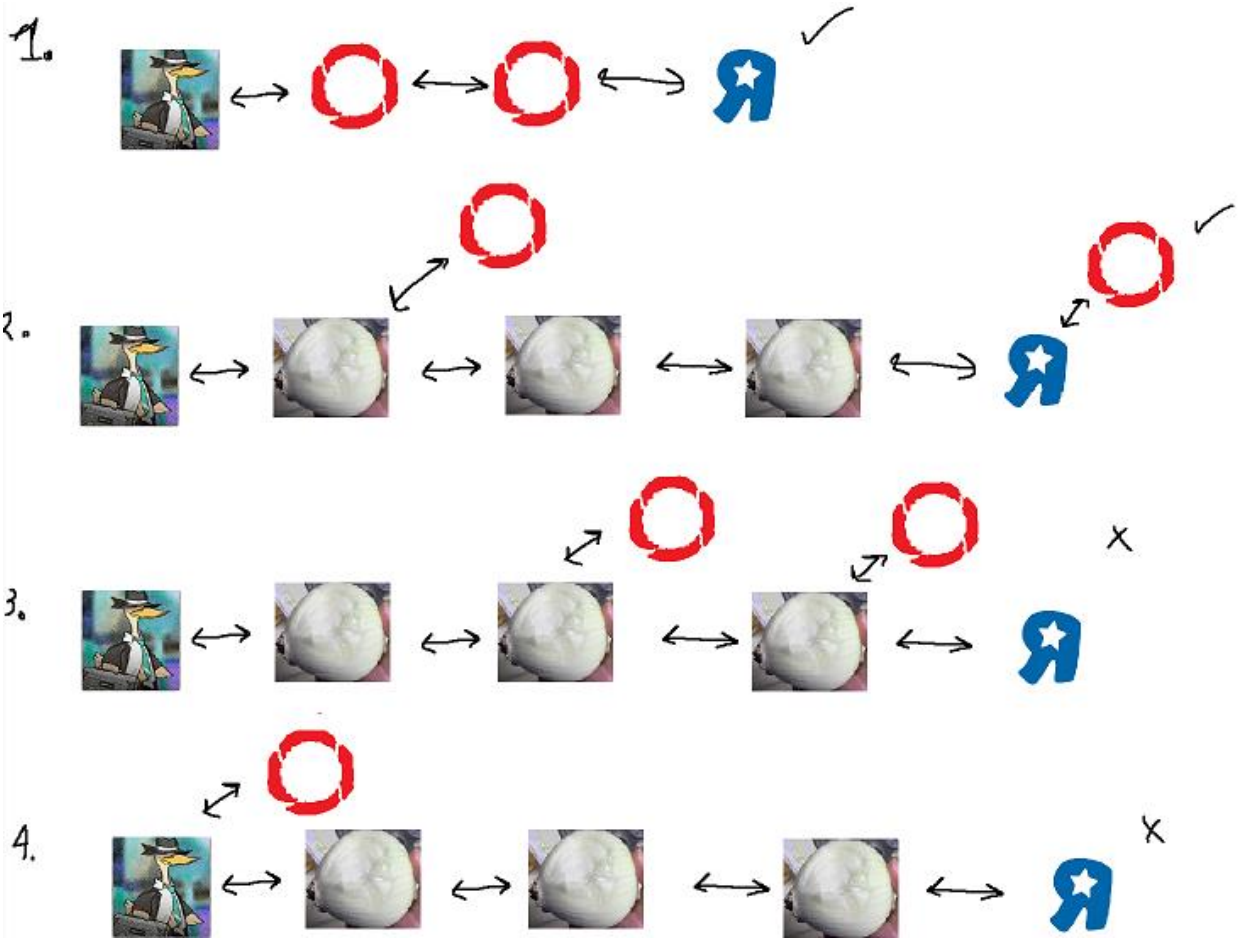
b. Alice sends purchase request to Bob, Mallory intercepts these messages and signs them as Bob. encryption/decryption key-pair **e_**, **d_** and signature/verification key-pair **s_**, **v_** where **_** is either A, B denoting Alice's, Bob's keypair respectively

i. Alice encrypts message to Mallory with **eA** signed with **sA**, Mallory decrypts with **dB** and verifies that the message is signed by Alice with **vA**

ii. Mallory encrypts message to Alice with **eB** signed with **sB**, Alice decrypts with **dA** and verifies that it is signed by Mallory using **vB**

Q3:

1. We can determine. We will be able to tell that Mr. Goose is connected to Streams “R” Us because we have access to the packet logs which will show packets originating from the streaming site with Mr. Goose’s desktop as the destination.
2. We can determine. Because we control the traffic at the entry node and the exit node (here we actually have control of the target website) of the tor circuit, we can perform a traffic correlation attack to identify that traffic from Mr. Goose enters the Onion circuit and a correlating traffic on the target website is observed soon after. So, we can reason that Mr. Goose’s is using Tor to gain access to the website.
3. We cannot determine. Access to the middle node is meaningless because we cannot see which node the traffic originated from and where it will end up so we won’t gain any information to determine if it is from Mr. Goose, similarly having access to the exit node we only know that someone is connected but not who it is.
4. We cannot determine if Mr. Goose is connected to Streams “R” Us because Mr. Goose will be connected to a Tor Onion router and that router will be connected to another, etc. and the final onion router will be connected to Streams “R” Us but due to tunneling between the onions we will not be able to see who is connected to the final Onion router.



Q4:

Full View

executor		Destination		Location Code		Mission		Salary		TC
Alice	C	Cryptomania HQ	TS	N2L 233	C	Spying	TS	5000	C	TS
Eve	C	East Lake Center	C	W1E 514	C	Marketing	C	1000	C	C
Charlie	C	Informatica HQ	C	123 G7A	C	Spying	C	5000	C	C

Bob's View (Filtering the table for users having Secret clearance)

executor		Destination		Location Code		Mission		Salary		TC
Alice	C	-	TS	N2L 233	C	-	TS	5000	C	TS
Eve	C	East Lake Center	C	W1E 514	C	Marketing	C	1000	C	C
Charlie	C	Informatica HQ	C	123 G7A	C	Spying	C	5000	C	C

1. Bob can infer from the information that is available to him that the location code which are zip codes will reveal the destination because they are mapped to a location. He has learned that Alice is executing a mission at Cryptomania HQ. Also, he can infer that because the salary for Alice's mission is the same as Charlies and Charlies mission was to a competing HQ and Alices mission is to a competing HQ that she is also on a spying mission. Although Bob has a limited view of the table he can manage to figure out the rest of the table entries.
2. To increase privacy we will use k-anonymity and t-closeness.

k-anonymity means that quasi-identifiers (in our case the zip code) should form an equivalence class to hide unique values and thus increase privacy. By reducing granularity such as hiding the location code's last 3 digits we will create an equivalence class on the first 3 digits of the zip code which generalizes to a much larger area rather than pointing to a specific building as it was previously. This is a tradeoff between privacy and utility. By hiding some of the executor's location code we are removing some information but there is still utility available to a data-analyst. We should also hide or altogether remove the names from the database otherwise the data will most likely not be in an equivalence class based on the zip code.

Even if we try to group the Salary into ranges it is still vulnerable to a skewness attack as higher salary missions will be of greater importance, and it will become easy to differentiate between marketing and espionage. This is when we introduce t-closeness. We can calculate a new salary for each equi-class (based on zip-code) using some sort of distance calculation that will change the distribution of salary. This will increase privacy but lose utility as the information is less exact.

1. The sensitivity is the value of how much the data can be affected by any single user, in this case a change in the number of matches a user has can affect the histogram by lowering a bin by 1 and increasing a bin by 1. This means that $\ell_1 = 2$
2. $\epsilon = 0.01, \frac{\ell_1}{\epsilon} = \lambda = 200$
3. Note that $d = 36$, $o_i = c_i + L_i$, where $L_i = Lap(200)$ and the variance of $Lap(200)$ is $2\lambda^2$

$$\mathcal{E} = \sum_{i=1}^{d} \mathbb{E}[(o_i - c_i)^2] = \sum_{i=1}^{d} \mathbb{E}\left[\left((c_i + L_i) - c_i\right)^2\right] = \sum_{i=1}^{d} \mathbb{E}\left[L_i^2\right] = 36 \cdot 2\lambda^2 = 2880000$$

4. Since we used the Laplace mechanism as our transformation, and we proved that the Laplace mechanism is ϵ-DP in lecture we can conclude that the mechanism does satisfy ϵ-DP. The histogram output of the mechanism will be rendered useless as the error is very large because the data will have a very large randomly sampled number added to it.
5. The algorithm will take as input $l \in \sum$ and output $o \in \sum$ by using the fraction of users in that location as the chance of sending that location as output. So if there are a total of 100 users and 15 users in l_1, 15 in l_2, 30 in l_3 and 40 in l_4 there will be a 15% chance for $o = l_1$, 15% chance for $o = l_2$, 30% for $o = l_3$ and 40% for $o = l_4$. This will result in an ϵ-LDP because it is a mechanism with a discrete input and output space, meaning we can compute an ϵ. We will need to calculate the probability that a user outputs a location given that they are from a location for all permutations of locations. An example is given below. Then we can compute ϵ by finding the maximum chance in each column divided by the minimum chance in each column, the maximum of these quotients as the input of the natural log will give us the ϵ privacy factor. Let $\Pr(l_0|l_0)$ be the global max and $\Pr(l_0|l_1)$ be the global minimum then we can calculate $\epsilon = \ln\left(\frac{\Pr(l_0|l_0)}{\Pr(l_1|l_1)}\right)$.

	$o = l_0$	$o = l_1$	...	$o = l_n$
$l = l_0$	$\Pr(l_0\|l_0)$	$\Pr(l_1\|l_0)$	...	$\Pr(l_n\|l_0)$
$l = l_1$	$\Pr(l_0\|l_1)$	$\Pr(l_1\|l_1)$	...	$\Pr(l_n\|l_1)$
...	...	...	...	...
$l = l_n$	$\Pr(l_0\|l_n)$	$\Pr(l_1\|l_0)$	...	$\Pr(l_n\|l_n)$