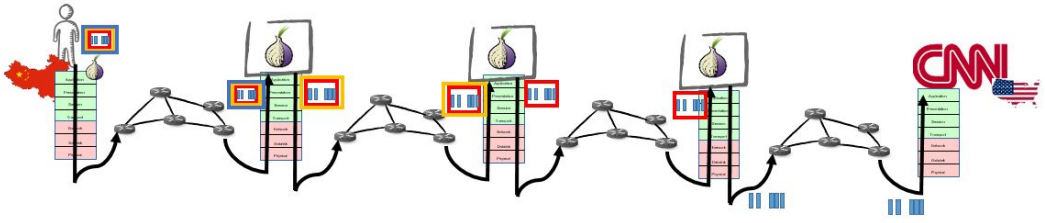
Exercises - Week 12
Privacy

1. Are the following statements True or False:

a) **Privacy technologies based on using access control to control who access what, protect you from the service provider**
b) **Anonymous communications based on onion routing only protect from adversaries that cannot see both ends of the anonymous communication networks**
c) **Encryption is the only protection needed to ensure privacy online**
d) **Anonymous communications are the only protection needed to ensure privacy online**
e) **An internet without routing security (e.g., enabling BGP hijacking) can break the protection given by low-latency anonymous communication systems**
f) **Using a VPN provides privacy from a local adversary that can only observe the connection from the user to the proxy but not from a global adversary that can observe any link in the network**

a) False, access control is implemented by the service provider, who enforces some access control policy. To implement this access control the provider does have access to the data, ie., there is no privacy towards the provider.
b) True, onion routing is a technique for low-latency anonymous communications networks. As such, it preserves patterns which can be recognized if an adversary can see traffic coming into the network and leaving the network.
c) False, encryption only protects content. Privacy online requires protecting also meta-data (e.g., identities, location, etc).
d) False, anonymous communications only provides protection at the network layer. To obtain full privacy one also needs to take care of the application layer (e.g., use encryption).
e) True. Anonymous communication systems are **overlay** networks. This means that onion routers are not network routers. They run at the application layer. In reality, a realistic picture of Tor looks like this:



Therefor by using BGP hijacking an adversary can reroute traffic in such a way that they can see both traffic coming in and leaving the anonymous communication network . As such they can put themselves in the position of launching an attack.[1] (think of the example seen in the class were all traffic from US was routed through Belarus. In this case a Belarus ISP would be a Global adversary for a US citizen visiting a US website would have an adversary)

f) True, a local adversary can only see the IP of the VPN service. Thus it cannot learn the destination of the traffic. However, a global adversary that can observe any link can see the input and the output to the VPN. A simple timing attack would suffice to track where incoming flows go.

2. We have seen in the class that a problem with the Tor network, and in general any low-latency anonymous communication systems, is that the shape of the traffic enables an adversary to trace flows through the network. A possible option to fight this type of attacks is to "pad" the connection, i.e., insert enough dummy packets so that all connections look similar. Does this approach make sense to:

a) Conceal whether a user is visiting a news website (e.g., CNN, BBC, etc) or a small static web such as http://www.guimp.com/

b) Conceal which EPFL employee website a user is visiting (e.g., https://people.epfl.ch/)

c) Conceal which movie you are downloading from the Pirate Bay.

d) Conceal when electronic votes are being sent.

[HINT: think about the bandwidth overhead you would need to protect the traffic]

a) Not really. Nowadays news websites contain a fair amount of images, videos, etc. As such, they require to download a fair amount of bytes. In order to conceal the small static webs, one would need to add enough "padding" for the small websites to look like the news websites. This creates a lot of overhead in the network.

b) Yes, it would make sense. In the case of the EPFL employees websites, all of them are very similar. Thus, very little padding would make them indistinguishable.

c) Not really. Similarly to the news websites, movies can be quite large and there may be a large difference between some of the archives in the database (from Megas in a small videoclip, to Gb in an HQ movie). The overhead to make them all similar would be quite large.

d) Yes, it would make sense. Votes are very small. With little padding, one could fake the sending of "dummy" votes effectively concealing when real votes are being sent.

[1] http://raptor.princeton.edu/

3. Let us assume that a COM-301 TA decides to invest the money from selling the exam questions into helping the Tor network instead of going to Hawaii. Thus, he buys 1000 machines and sets them up in the basement as onion routers.

(a) Does this increase the Tor network capabilities to offer anonymity? Against which adversary?
(b) If instead only one assistant, each of the COM-301 students puts 10 machines in their own basement, would the situation improve?

[HINT: think about who could see the traffic going through the new machines should they be incorporated to the Tor network.]

a) No, it does not. Having many servers increases the number of possible routes, but if all of them are in the same place there is no diversity. An adversary that can see one machine can see all of them (i.e., can see the full Tor circuit and use traffic analysis to find where flows go). The problem of this approach is the little diversity in the nodes that puts certain adversaries in an advantageous position. In this case, for instance the Swiss ISP serving the TA's basement internet.
   Note that these additional servers do not increase anonymity with respect to an adversary that cannot see them either. For this adversary, it actually makes no difference that there are more or less circuits that can go through the TA's basement. On the other hand, having more nodes does increase the capacity of the network.
b) This improves the situation, for instance against the ISP if the students are served by different ISPs. Yet, for instance the Swiss government could have the capability to observe the traffic in all of them (e.g., with a subpoena).

4. Let us assume you are a service provider designing a new recommendation system for best restaurants in campus. Assume a simplified environment in which there are three actors: the students using the application, the restaurant owners, and the service provider serving the application.

Compare the following configurations in terms of privacy (i.e., privacy risks with respect to other entities in the system) from the point of view of the students.

CONFIG A: The application gathers the recommendations from the students and then: lets other students see each other recommendations, and lets the restaurants see the student recommendations so that they can offer discounts to students that give good ratings.

CONFIG B: The application gathers the recommendations from the students and then: lets other students and the restaurant owners see the average rating for a restaurant.

CONFIG C: The students application compute a ranking of the restaurants and uses advanced cryptography to send this ranking to the service provider. This cryptography enables the service provider to compute a global ranking without seeing each individual student opinion. The restaurant owners receive the global rating.

First, think that having access to the ratings reveals when and where a student has had lunch. Depending on the type of restaurant this may reveal further information. For instance the student is vegan, the religious orientation of the student (e.g., restaurants with no pork, restaurants offering kosher food), the health condition of the student (restaurant specialized in food without gluten), etc.

a) This configuration is very bad for privacy. In this configuration other students, restaurant and service provider see all the recomendation and ratings.
b) This configuration is better, but the service providers still sees all the data. Others only see aggregates.
c) Best option, only the student herself gets access to her own ratings.

QUESTIONS ABOUT PREVIOUS TOPICS

5. When a client wants to connect to a web server with TLS, it needs to know the public key of the server. However, a client cannot know the key of every site on the internet. Public Key Infrastructure (PKI) was created to solve this problem: it provides clients with the public keys of the servers they want to communicate within a trusted manner. The nodes which provide public keys are called certificate authorities (CA).

a) How can we ensure the integrity of the message that we receive from a CA?
b) How can a user get a list of trusted CAs for the first time?
c) Is having only one CA enough for the internet? What about thousands of CAs?
d) What is the best structure for CAs?
e) Should public keys be permanent? If not, when should they get replaced?

a) The CA signs every owner/public key relation which it validates. The signature of a CA on the ownership of a public key is called a certificate. However, this requires the client to know the public key of the CA.
b) Operating systems and browsers come with an internal set of CA certificates.
c) Having only one CA would create a single point of failure for the internet and the load on this CA would be too much to handle. Furthermore, everyone needs to trust this one node and considering the technical and political problems in the world, having one entity which is trusted by everyone (USA, China, Europe, Russia, Iran, North Korea ...) is infeasible. Replicating this CA to thousands of nodes would solve the functional single point of failure and the load handling, but the adversary only needs to compromise one of the thousands CA to be able to steal the identity of anyone.

d) Hierarchy is one of the common solutions for this problem. There are dozens of root CAs as the core of the PKI which are hardcoded into OSs and browsers, and every other CA gets a direct or indirect credential from higher level CAs.
e) No, every certificate has an expiration date. Having longer expiration date gives adversaries more time to break or steal the private key, on the other hand, has a short lifetime, requires frequently issuing and asking for new keys. Hence, the lifetime of the key gives a tradeoff and the owner need to consider, the security, impact of leakage, and frequency of use of a key before deciding on the expiration date. Moreover, private keys get stolen or lost and we need to have a way to dynamically revoke certificates. CRL and OCSP are two major protocols which handle certificate revocation.

6. Analyze the following message combinations for the following properties.:

1) **Provide confidentiality against a MITM adversary**
2) **Provide integrity against a MITM adversary**
3) **Allow Bob to read the message**

(a) Alice sends to Bob: Enc(PK_{Alice}, M), Sign(SK_{Alice}, SHA(M))
(b) Alice sends to Bob: Enc(k, M), Enc(PK_{Bob}, k)
(c) Alice sends to Bob: Enc(k, M), Enc(PK_{Bob}, k), Sign(SK_{Alice}, SHA(M))
(d) IV, Enc(PK_{Bob}, k), M $\oplus$ $STREAM_k$ $\oplus$ IV

PK_{Alice}= Public key of Alice
SK_{Alice}= Secret signing key of Alice
PK_{Bob}= Public key of Bob
Enc(pk,data) = Public-key encryption of data with the key pk
SHA(M) = SHA-256 hash of message M
$STREAM_k$ = output of a stream cipher with key k
k = symmetric key
M = message

a) This protocol provides integrity because it signs the message with Alice's secret key. Nobody can read the message except Alice. The message is encrypted for Alice and only she can decrypt it. However, the message is encrypted with a asymmetric key, so it must be very small.
b) In this protocol symmetric is encrypted for Bob, so only Bob can read the message. There is no integrity check, so the MITM can easily change everything.
c) Similar to b only Bob can read the message, but the MITM can no longer change the message. Alice signs every message with her secret key, and she is the only person who can sign under her name.
Here, the IV and $STREAM_k$ act as a stream cipher and they are equivalent to Enc(k, M) and this protocol has similar properties to b.