5. Today Bitcoin is very popular and its use is increasing rapidly. Discuss potential problems that might arise in the future and that might dramatically limit the use of Bitcoin.

   A miner has incentives if the total costs (hardware + power) are less than the total reward, i.e. the block reward and the sum of transactions fees. The consensus protocol is based on the fact that miners will act as good nodes in the network and this can happen only if they are interested in being good, i.e. they earn something. The necessary hardware to be a miner is increasing its cost due to the fact that the hash power needed to solve the puzzle to propose a valid block is increasing as well. From the cost point of view, small miners are kicked out from the network, this means that only big miners company have the necessary capabilities to propose blocks. Less diversity in miners means less decentralization in bitcoin network. Moreover the blockchain will be extremely large, due to the always increasing number of transactions. This means that a miner needs a really efficient and expensive hardware to implement validity checks in order to avoid double spending and other malicious activities from network's users. Notice that also the block reward is halved after a fixed number of proposed blocks, this is because the block reward follows a geometric series and it has a maximum final number of possible bitcoins in the network. This situation is still not well studied from an economic point of view. There will be the necessity of always increasing transaction fees when the block reward will reach the zero value, in order to let miners to be interested in acting as a good node. Finally, the bitcoin software is hard to update due to hard/soft forks problems. This brings the risk of not up-to-date situation. A software not updated will loose credibility from users and this could bring to the end of bitcoin network. Credibility is also affected by bitcoin reputation attacked by malicious users. A malicious not controlled users can fake some transactions, a miner will not block him if it is not acting as a good node.

6. Illustrate Mixing in Bitcoin discussing its advantages and its limitations; discuss digital coins that allow anonymity.

   Bitcoin activities are recorded and available publicly via the blockchain. It means that a third party can trace your transactions and find ID information. Mixing services are used to mix one's funds with other people's money, intending to confuse the trail back to the funds' original source. When mixing bitcoins, you send your money to an anonymous service and they will send you someone else's coins. The problem with a dedicated mixing service is that we must trust it: it promise not to keep records, it doesn't ask for your identity, it must only swap user's coin with other users' coins. In order to increase anonymity (better increase the difficulty of tracking back some transactions) mixing services can use multiple mixing rounds, e.g my money will be gave to an user B and the ones of the user C (in another mixing round) will be gave to me. In order to increase anonymity also the transactions must have all the same bitcoin value! If trusting in a third party service is too dangerous there is a decentralized mixing scheme that provides the same kind of service but requires some more pre-activity. It is called Coinjoin and exploits the join payment mechanism provided by bitcoin itself: first of all we must find peers that want to be involved in a mixing transaction, then we exchange input/output information, we construct a single transaction with multiple input and multiple output, we all sign the transaction in order to be compliant to bitcoin and then broadcast the transaction waiting for its insertion in the blockchain from a miner. This approach provides a better anonymity since no third party entity can log anything but of course it suffers of many problem related to the decentralization scheme like how to find peers and only the involved peers know each other identity and the outcome (which it is worst wrt to the third party entity, one of the peer could be a malicious one and can publish transaction's information).

One digital coin that provides anonymity is zerocoin. First of all, we can say that Zerocoin brings mixing to protocol level. So, inserting it into the protocol, it gives a cryptographic guarantee of mixing. In fact, is no more necessary to trust a third party mix or a set of mixes. But, you would only need to rely on the underlying cryptography being solid. We can see Zerocoin as a cryptographic proof that someone owned a Basecoin and that it was made unspendable. And miners will be able to verify this proof. That's what gives the right to later redeem a new Basecoin in exchange for the Zerocoin. The used proof is called a zero-knowledge proof that gives the possibility to someone to prove a statement without reveling any information about the statement itself. In order to generate a zerocoin an user needs to: generate a random serial number S of the zerocoin, another random number r to be kept secret, compute H(S,r) and put this commitment in the blockchain spending some Basecoin, finally, in order to redeem a different Basecoin with the same value, prove to the miners that you own the value r and you previously claimed some Zerocoin. This allows you to obtain a total different Basecoin from anyone previously put in the blockchain.

7. What is pseudo anonymity and discuss key features with reference to a practical example. Explain the advantages and weakness of pseudo anonymity discussing how it should be used in practical situations. Or Pseudo anonymization is an important technical element in the GDPR. Present main ideas of pseudo anonymization, its advantages and its limitations.

   Pseudonymity is the use of a false name or something that represents an entity without reveling some sensitive information that allow others to figure out the real identity of the entity. A practical example is the pseudonyms used in Bitcoin network: each user who wants to join the network is associated to one or more pairs of public and private keys, in order to refer to some user we must use the hash of one of his public keys as its wallets' address. In this case an user is not asked to insert any sensitive information, however it is assigned to him a pseudonym, i.e. the public key itself. The advantages are mainly related to the possibility of hiding the real identity of any users: in order to be more specific we are also allowed to create multiple users related to our identity, without a limit, since there is no limitation related to our real identity. Each one can act under an unknown unlikable pseudonym. However this is only in theory: the main problem about pseudonym is that it doesn't hide any relations between performed actions of the same user, they will be linked to the same pseudonym. This is also the case of Bitcoin: since the blockchain, i.e. the ledger, is public, then we can inspect it figuring out and linking all the previous transactions performed by the same pseudonym. We can, in principle, build a transaction graph, inserting all the pseudonym involved, and through external sources of information, try to identify some pseudonym searching for his real identity. So, in a nutshell, pseudonym hides the real identity of a user allowing him to act under a fake name or an unlikable information but doesn't protect our actions within the service to be observed and linked together in order to re-build our usage history of the service itself.

8. Discuss how differential privacy has been used in the US census. In particular discuss an example that motivates why publishing census data does not preserve privacy and discuss the kind of operation that are performed to achieve differential privacy.

   The 2020 Census brings a new era of privacy protections with the implementation of differential privacy. Differential privacy is a "formal privacy" approach that provides proven mathematical privacy assurances by adding uncertainty or "noise" to the released data. This disclosure avoidance technique determines the amount of noise necessary to balance privacy loss and accuracy via mathematical formulas. In a nutshell: withholding or changing one person's data must not substantially change the algorithm's output. They decided to compute and publish histograms into six level of degrees: from a

national level to a block level of measurements (the block is the smallest geo unit of the census). They designed a new algorithm scheme called top-down mechanism: first of all the algorithm produces the national level histogram without any geographic identifiers, then, proceeding from nation to the smallest block level, it computes the different noises to add at each level such that the total sum of the noises must be equal to the one added to the national histogram. This was done in order to prevent inconsistency of data between different level of measurements. This approach is easy to parallelize and reduces variance for many aggregate regions. These partial results are then processed by two optimizer: the first one work on child histograms in order to add up to parent ones, the second one maintain consistency within the children and the national one. This approach was proved to be much more accurate than the ones than in the past years (higher level of measurements higher the accuracy).

Achieving differential privacy is a really delicate and precise work to perform. It needs a lot of analysis of the data that we want to publish in order to understand what are the dangers and the possible leaks that could happen in our specific case. Although the differential privacy definition could seem easy to understand, it is not so easy to implement in practice: the noise that must be chosen is the output of a trade off between privacy and utility/accuracy of the data. More noise means more privacy and less accuracy, the inverse holds too. This means that whenever we decide a particular value of the noise ϵ, we are implicitly asserting the measure of leaked information of the published data. This parameter is decided without knowing what are the information that an attacker has collected from external sources, i.e. what are the information that he will exploit for identifying someone that participated to the census. For example let's imagine that for a particular category of interest like black teenager women of an high school, just one woman reaches the highest result at the end of its educational career. If the information is published as it is, an attacker could inspect the annuals of the school in order to retrieve the identity of that woman. So a noise needs to be add in order to level that result to the mean/median of the other categories. But how useful is the modified information? It is meaningful publishing data for a so small poll's population? This is the trade off between privacy and accuracy that I mentioned before.

9. What are Data Protection Impact Assessment (DPIA) goals and for which reason is important.

   Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. E.g. if you are using new technologies or tracking people's location or behavior or processing personal data related to sensitive information. You must prepare your DPIA before beginning any data processing activity. Ideally, you should conduct your DPIA before and during the planning stages of your new project. It is a systematic description of the required processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller. It is an assessment of the necessity and proportionality of the processing operations in relation to the purposes. It is an assessment of the risks to the rights and freedoms of data subjects. It is important because is one of the most important ways to demonstrate to authorities that your organization complies with the GDPR.

10. Privacy by Design (PbD) is an important technical element in the GDPR. Present main ideas of PbD, its advantages and its limitations.

    The most basic explanation of Privacy by Design is little more than "**data protection through technology design**." At its core, it means that you need to integrate data protection and privacy

features into your system engineering, practices and procedures. It shouldn't be an afterthought or a supplement to your processes or infrastructure. There are seven principles in the concept of Privacy by Design and each one is just as important as the next. These principles are:

1. Proactive not Reactive/Preventative not Remedial
2. Privacy as the Default
3. Privacy Embedded into Design
4. Full Functionality
5. End-to-End Security
6. Visibility and Transparency
7. Respect for User Privacy

The first principle argues that data privacy needs to come up at the beginning of the planning process. The second one means that privacy needs to be at the forefront of what you do, i.e. using data minimization, deleting data you no longer use and so on so forth. The third one means that privacy is a core functionality of the product. You should be using encryption, authentication, and testing vulnerabilities on a regular basis. Fourth one tells that no functionalities must be sacrificed for privacy, it must be accomplished within every kind of service. The fifth one means that privacy protection follows data through its lifecycle, from collecting it to deletion/archival. The sixth one follows the principle no security by obscurity, users must be aware of privacy and how it is managed. The last one means that everything needs to remain user-centric, i.e. the data must belong to the consumer you collected it from.

However, the legislation doesn't name the exact measures to be taken beyond features. Indeed, Privacy by Design is now considered only to be a best practice for all organizations that engage in data processing, no matter how big or small. Each data processing must be managed following those principles, but the real techniques are up to the companies. Indeed, they must shows that they're following them with other documentations like DPIA.