2. We have four data privacy techniques:

   1. k-anonymity
   2. Differential privacy
   3. Secure multiparty computation (SMPC)
   4. Private information retrieval (PIR)

   For each scenario below, two of the four data privacy techniques will be proposed to resolve the challenge. Choose the correct one, explain why it is suitable, and why the other choice is not suitable.

   (a) [4 points] You want to buy a new smart device that encourages a healthy lifestyle by monitoring your daily exercise. The device needs to track your movement on a map to know how much calories you are actually expending (e.g. hiking and swimming is different from walking). However, you consider this to be a privacy risk: you do not want the app to know where you are at all times. A new company making these smart devices is willing to use a data privacy technique to protect your privacy. Proposals: differential privacy, PIR.

   (b) [4 points] You would like to purchase a new web domain. However, you are aware of the practice of cybersquatting; people may purchase the domain first if they know it is in demand, and sell it to you at an elevated price. You want to know if the domain is still available, but you are worried that attempting a DNS query for the domain will lead to some DNS servers purchasing it immediately for cybersquatting. To assure potential customers that it is not malicious, a DNS server is willing to cooperate with you and implement a privacy-preserving algorithm. Proposals: k-anonymity, PIR.

   (c) [4 points] A hospital has information about millions of carcinogenic and non-carcinogenic patients in various locations. The exact location, if revealed, is a privacy breach. Researchers want to determine if there is a correlation between cancer and building materials; they already have a large database of building materials used in different areas. The correlation is potentially complicated, though constructions in the same area almost always use the same building materials. Proposals: k-anonymity, SMPC.

   (d) [4 points] A new service has been developed to allow users to compare salaries and potentially identify unfairness or discrimination in pay. After passing a local verification process, users can submit their salaries to the service, and the service should be able to inform them if they are being underpaid or not by comparing with other users of the same profession/experience (once there is enough data). The actual salary is considered sensitive information and should not be exposed to the service. Proposals: differential privacy, SMPC.

Solution:

Partial credit of 2 points is possible for choosing the wrong one if a true disadvantage is given for the correct choice. For example, SMPC is slow, k-anonymity has issues with attacker background knowledge, differential privacy is noisy, PIR is also slow. 1 to 2 points if choice is correct but reasoning is wrong or insufficient.

(a)
- Correct: PIR can be used to fetch the calorie information for different activities to different locations from the server without revealing your location.
- Incorrect: differential privacy will add noise to your location, which can lead to significant errors because the app may think you are swimming in a river when you are actually running on a road.

(b)
- Correct: PIR can be used to fetch the status of domain names without revealing the domain name.
- Incorrect: k-anonymity will add undesirable noise to this problem/there is no QID to anonymize.

(c)
- Correct: k-anonymity can be used to anonymize the address of patients before publishing it for researchers.
- Incorrect: SMPC is not practical for huge data sets/when potential algorithms are complicated.

(d)
- Correct: SMPC can be used to achieve this; the querier inputs their salary, the service inputs other users' salary, the algorithm returns if the querier's salary is lower or higher.
- Incorrect: differential privacy adds a large amount of noise to one person's data. This means that the querier will not be able to obtain useful results.