CS 523 Final Exam, 14.08.2020

Name: **Chris P. Chicken**

Sciper: **123123**

Please wait for instructions before opening this document

- You can bring with you an “aide-mèmoire" of a most 3 A4 pages (not 3 recto-verso sheets).
- You are not allowed to bring a lens.
- You are not allowed to use electronic devices during the exam.

Multiple choice questions:

- There are 20 multiple choice questions, each worth 1 point.
- Only one answer is correct; *there is a 0.25 point penalty for wrong answers*
- Make a mark *inside* the box corresponding to your answer
- Use a black or blue pen to mark your answers. Pencils are not allowed.
- Use white-out fluid or tape if you ticked the wrong answer.
- If you white-out a wrong answer, do not try to re-draw the boxes.

Open text questions:

- There are 11 open text questions, each worth 2 points.
- Please write your answers in the corresponding text boxes.
- Use a black or blue pen to mark your answers. Pencils are not allowed.
- Do not write more than the lines specified in the box. Any text outside of the boxes **will be ignored**.
- Do not tick the grading boxes of the top of the text boxes.
- Please mind your calligraphy; undecipherable responses will not be graded.

Questions

- The supervisors will not answer any questions regarding the content of the exam questions

Part 1: Multiple-choice questions
Mark the correct answer by *completely* filling in the corresponding square (■).
There is **only one** correct answer per question.
Each correct answer earns 1 point. Incorrect answers have a penalty of 0.25 points each. No answer neither adds nor subtracts points. Multiple marked answers are considered incorrect and will result in 0.25 points reduction.
Use white-out fluid to change (delete) an answer (other deletion methods yield 0.25 reduction).
Read the answers carefully; the template may split the answers across columns.

Question 1 **[Interpersonal Privacy]** Which of the following statements is *TRUE*?

- ☐ Co-location is an efficient privacy-preserving mechanism.
- ☐ Not sharing my personal information prevents co-location attacks on my location privacy.
- ☐ Co-location attacks on location privacy never need background information.
- ☐ Co-location information can increase the effectiveness of attacks on location privacy.

Question 2 **[Beaver]** Which of the following statements is *TRUE* ?

- ☐ Beaver triplets are a solution to avoid the need for a trusted setup in multi-party computations.
- ☐ Beaver triplets enable non interactive multiplication.
- ☐ The only way to generate Beaver triples is to delegate this task to a third party
- ☐ Beaver triplets enable the multiplication of two shared values for a finite number of parties.

Question 3 **[RLWE]** Which of the following statements about BGV is *FALSE*?

- ☐ A BGV noise which is too big may compromise the correct decoding of the ciphertexts.
- ☐ The norm of the BGV noise grows exponentially with the number of additions.
- ☐ BGV is a homomorphic encryption scheme whose security builds on the hardness of the ring learning with error (RLWE) problem.
- ☐ Given a ciphertext encrypted under BGV scheme, finding the secret key is as hard as solving the search RLWE problem.

Key	Gender	Zipcode	Age	Disease
Eric	M	1007	25	Cancer
Justine	F	1012	25	Heart Disease
Emma	F	10**	25	Flu
Helen	F	1012	*	Flu
Paul	M	1007	25	Cancer
Philip	M	1012	35	Herpes
Michel	M	1012	35	Cancer
Mory	M	1007	25	Cancer
Adrien	M	100*	25	Heart Disease
Mallory	M	10**	35	Flu
Camille	F	10**	25	Herpes
Samuel	M	1012	35	Cancer
Marco	M	1007	*	Cancer
Damien	M	10**	35	Flu

Table 1: Hospital database: The *Key* column represents patients. The *Gender*, *Zipcode*, and *Age* are patient's attributes. The "*" symbol can take any value.

Question 4 **[k-anonymity]** Consider only the *Gender*, *Zipcode*, *Age* attributes in Table 1. Which statement is *TRUE*?

- ☐ The database achieves k-anonymity with $k = 4$.
- ☐ The database does not achieve k-anonymity for any k.
- ☐ The database achieves k-anonymity with $k = 1$.
- ☐ The database achieves k-anonymity with $k = 2$.

Question 5 **[l-diversity]** Consider *age* and *disease* to be sensitive attributes in Table 1. Which statement is *TRUE*?

- ☐ The database achieves 3-diversity.
- ☐ The database is differentially private.
- ☐ The database achieves 5-diversity.
- ☐ None of the other answers.

Question 6 **[Differential Privacy]** Consider a database holding electricity consumption of 100 households. The power company wishes to publish the average consumption of the neighbourhood in a privacy-preserving manner. To this end, they employ differential privacy with a Laplace mechanism. Remember that this implies perturbing the result with noise sampled from the Laplace distribution with a scale parameter b. We note Δf the sensitivity and ε the privacy parameter. Which of the following statements is *TRUE*?

- ☐ The larger the range of the power consumption values is, the larger the sensitivity is.
- ☐ The sensitivity of the query does not depend on the database.
- ☐ The scale of the Laplace distribution is $b = \frac{\varepsilon}{\Delta f}$.
- ☐ Bigger ε implies better privacy.

Question 7 **[ABCs]** Which of the following is *NOT* a property of attribute based credentials (ABCs)?

- ☐ Only the issuer is able to provide valid credentials.
- ☐ The verifier should not be able to link two consecutive showings of the same credential.
- ☐ The issuer should keep track of the ABCs it is issuing.
- ☐ The prover can select which of her attributes to reveal to the verifier.

Question 8 **[Unobservability]** Which system provides unobservability against a global, passive adversary?

- ☐ Mix-Net.
- ☐ Crowds.
- ☐ DC-Net.
- ☐ Tor.

Question 9 **[Modular Arithmetic]** Consider $\mathcal{R}_5 = \mathbb{Z}_5/(X^2+1)$. Let $P(X) = 3X^2 + 2X + 1$ and $Q(X) = 7X - 1$. Which of the following is $P(X) \cdot Q(X)$ in $\mathcal{R}_5$?

- ☐ $21X^3 + 41X^2 + 25X + 8$
- ☐ $4X + 3$
- ☐ $4X - 3$
- ☐ $X^2 + 3$

Question 10 **[Steganography and Watermarking]** Which of the following statements is *TRUE*?

- ☐ Steganographic techniques can always achieve higher capacity than watermarking techniques.
- ☐ None of the other answers.
- ☐ Watermarking requires the embedding to be robust to intentional and non-intentional attacks.
- ☐ Watermarking focuses on concealing the covert communication channel.

Question 11 **[Online tracking]** Consider a user visiting `AwesomeWebsite.com`, which contains advertisements from `ILoveAds.com`. Which one of the following statements is *TRUE*?

- ☐ A cookie set by `ILoveAds.com` is not a third-party cookie since `ILoveAds.com`'s content is present on the website the user is visiting.
- ☐ Cookie syncing allows another website, `UserSpy.com`, to get information from a user's cookie, even if `UserSpy.com`'s content is not present on `AwesomeWebsite.com`.
- ☐ If the user visits a different website that also contains advertisements from `ILoveAds.com`, the user's sessions cannot be linked by `ILoveAds.com` via cookies. Only multiple visits to the same website can be linked.
- ☐ If the user uses the Safari browser, which disables third-party cookies by default, they are protected from any tracking since cookies can't be set.

Question 12 **[MIA]** Assume a non-trivial (better than a random) machine-learning classifier is trained in such a way that it satisfies differential privacy (DP) with parameter ε. Which of the following statements correctly characterizes the relationship between the DP property of the classifier and the success of membership inference attacks (MIAs) against this classifier?

- ☐ ε-DP does not impact the success of MIAs.
- ☐ As ε decreases, the chance of MIA success decreases.
- ☐ As ε increases, the chance of MIA success decreases.
- ☐ ε-DP prevents any attacks against privacy of the training data, including MIAs.

Question 13 **[Active censor]** A censor suspects that a user u is using Meek (domain fronting) to bypass censorship. The user frequently connects to the FreeCloud cloud provider. What is the best option for determining the possibility of having Meek connections with FreeCloud?

- ☐ Setting TCP reset in FreeCloud's packets to u.
- ☐ Delaying u's packets to FreeCloud.
- ☐ Re-routing u's connection to FreeCloud
- ☐ Directly probing FreeCloud

Question 14 **[Secret registration]** You have to design an approach to register users in the *SecretStroll* project. Which option is the best?
Consider a user who has subscribed to $S = \{s_1, s_2\}$ from the set of all subscriptions U.

- ☐ Hide: $\{\text{secret\_key}, S\}$, Reveal: $\{\}$
- ☐ Hide: $\{\text{secret\_key}, s_1, s_2\}$, Reveal: $\{\}$
- ☐ Hide: $\{\}$, Reveal: $\{\text{secret\_key}, s_1, s_2\}$
- ☐ Hide: $\{\text{secret\_key}\}$, Reveal: $\{S\}$

Question 15 **[Crowds]** Which one of the following statements about the Crowds anonymity network is *TRUE*?

- ☐ It cannot resist against an active adversary who can delay packets.
- ☐ Longer conversations are less likely to be de-anonymized.
- ☐ It can protect against globally passive adversary.
- ☐ The previous node in the chain is more likely to be the initiator.

Question 16 **[k-anonymity cloaking]** Consider a k-anonymity cloaking system, where users send their location and query to a trusted third-party location anonymization service (LAS). The LAS computes the cloak – an area based on the user's location that also contains k-1 other users. The LAS sends the cloak and the anonymized query to an untrusted location service provider (LSP). Which of the following statements is *TRUE*?

- ☐ Increasing the value of k always results in an improvement in users' location privacy.
- ☐ If the LSP is able to determine the location from which a user's query originated, the user's location privacy is impacted. However, the user still has k-anonymity.
- ☐ The location distribution of users in the system does not have an impact on their location privacy since cloaking prevents the LSP from knowing actual locations.
- ☐ k-anonymity cloaking provides location privacy even if the server tries to use statistical information about users' mobility patterns to infer their location.

Question 17 **[Cell ID fingerprinting]** In the SecretStroll project, you analyzed Tor traffic to perform cell fingerprinting. If the system would implement a countermeasure to prevent fingerprinting, which of the following options would be the least effective to stop the attack?

- ☐ The app batches queries for multiple cell IDs and changes their order every time before sending (assume that the app sends multiple queries in a day).
- ☐ The service provider adds dummy Points of Interest (POIs) data for every cell query such that the number and sizes of POI data sent in the responses from the service is always the same.
- ☐ The app sends a randomly chosen cell ID query along with its actual query.
- ☐ The service provider sends a random number of ACK packets, in addition to the query response, to each cell ID query by the app.

Question 18 **[Website Fingerprinting]** Which of the following features is the least useful for an adversary that mounts a website fingerprinting attack against Tor traffic?

- ☐ Average number of sequential outgoing packets.
- ☐ Average number of sequential incoming packets.
- ☐ Average Tor cell size.
- ☐ Total transmission size.

Question 19 **[Circuit Depth]** Which way of evaluating the following polynomial $p(x) = \sum_0^7 a_i x^i$ leads to the smallest multiplicative depth of the resulting circuit? (Note that intermediary values can be reused in the circuit)

- ☐ $a_0 + a_1x + (a_2 + a_3x)x^2 + (a_4 + a_5x + (a_6 + a_7x)x^2)x^2$
- ☐ $a_0 + a_1x + a_2x^2 + a_3x^3 + (a_4 + a_5x + a_6x^2 + a_7x^3)x^4$
- ☐ All the answers have the same multiplicative depth.
- ☐ $a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6 + a_7x^7$

Question 20 **[Location Privacy]** Consider a privacy-preserving location release mechanism that protects the users of SecretStroll against the service provider. Which of the following is *NOT* a measure of privacy that the mechanism provides?

- ☐ Expected distance from the released location to the true location.
- ☐ Precision of an attack aiming to identify the home location based on released locations.
- ☐ Recall of an attack aiming to identify the home location based on released locations.
- ☐ Expected distance from the released location to the points of interest.

Part2: Short answer questions: Write your answer using *only* the lines provided. Anything beyond the specified number of lines will not be considered for grading.
Answers are graded on a scale from 0 to 2.
Please mind your calligraphy; undecipherable responses will not be graded.

UMix

Alice and Bob want to secretly chat with each other without letting anyone know about their conversations. Alice has heard about a new mix-net anonymity system called UMix. UMix consists of three layers, and each layer only has one mix node. The system works in rounds. In each round, mixes wait for a fixed period of time, then shuffle the messages they have received and forward them to the next layer or to the receivers. Fig 1 shows an overview of UMix.

To thwart traffic analysis, in UMix participants send dummy messages to all other users. In each round, every participant s who does not have a message to send decides to send a dummy message with a probability of 5%. To send a dummy message, the sender chooses a random recipient r uniformly from the list of all participants except himself and sends r a message.

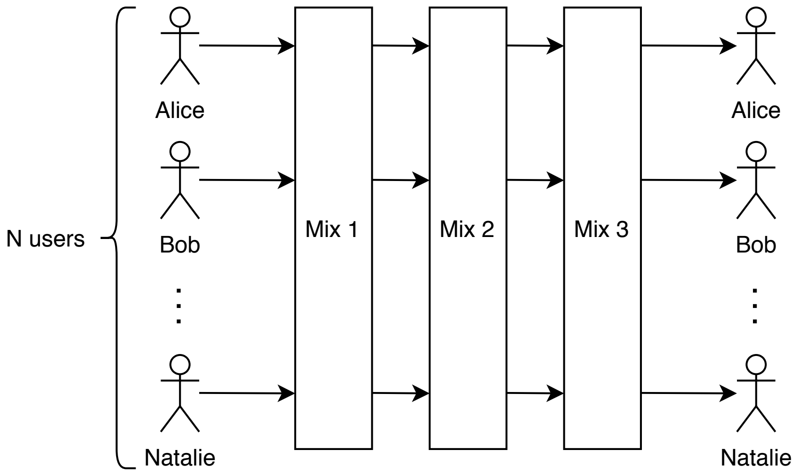

Figure 1: An overview of UMix

Question 21 Assuming that there are N active users ($N > 2$) in the system, and Alice and Bob do not chat with anyone else in UMix, can an adversary who gains control over mixes 1 and 3 detect whether Alice and Bob talk to each other? Justify.

☐ **0** ☐ **0.5** ☐ **1** ☐ **1.5** ☐ **2**

....

....

....

Censorship

Consider a region where a censor blocks the website `blocked.com`:

Question 22 Alice wants to access `blocked.com`. Tor traffic is blocked in the region. She considers two options:
a. Use a service that modifies her Tor traffic patterns to look like the traffic of a popular music streaming platform.
b. Use a service that utilizes emails as a carrier for her web traffic. She sends traffic via a public email provider (such as Gmail) to a proxy outside the censored region. The proxy forwards the traffic to `blocked.com`.
Name one disadvantage for each of these options. ☐0 ☐0.5 ☐1 ☐1.5 ☐2

....

....

....

Question 23 A censor wants to prevent that their user visits `blocked.com` based on traffic patterns. For this, it wants to use a machine-learning based detector. If the detector predicts that the user has visited `blocked.com`, it drops the user's connection. On average, 20% of all internet connections in the region visit `blocked.com`. Consider two detectors:
a. A detector that has 70% true positive rate (TPR) and 5% false-positive rate (FPR)
b. A detector that has 99.99% TPR and 50% FPR.
Which of the detectors would you use as censor that wants to minimize the accidental blocking of traffic to non-blocked websites? Justify numerically. ☐0 ☐0.5 ☐1 ☐1.5 ☐2

....

....

....

Contact Tracing

Question 24 The goal of epidemiological contact-tracing apps is to notify those who were physically exposed to people who have a specific infectious disease. Consider two hypothetical designs of such an app:

a. The app continuously broadcasts a user's unique identifier over Bluetooth to nearby smartphones. The app sends all "seen" identifiers (identifiers that are broadcasted by other people and are received by the user via Bluetooth), the GPS locations and timestamps of the encounters, and the user's own unique identifier to the central server.

b. Same as (a), but the app does *not* send the GPS locations.

Compare the two designs in terms of the *location privacy* and *social relationships privacy* with respect to the central server.

☐ 0 ☐ 0.5 ☐ 1 ☐ 1.5 ☐ 2

Legal

Question 25 Is obtaining user consent the only possible lawful basis to collect and process personal data? If so, describe how consent should be obtained. If not, describe other legal basis.

☐ 0 ☐ 0.5 ☐ 1 ☐ 1.5 ☐ 2

Vernam Cipher

Question 26 Prove that the Vernam cipher is malleable but not homomorphic with respect to the XOR operation.

We recall here the Vernam cipher. For a plaintext $m \in \{0,1\}^*$ randomly sample a key of same length $k \in \{0,1\}^*$. To encrypt, XOR the message with the key:

$$c = \text{Enc}_k(m) = m \bigoplus k$$

Similarly, the decryption follows the same protocol:

$$m' = \text{Dec}_k = c \bigoplus k$$

☐0 ☐0.5 ☐1 ☐1.5 ☐2

....

....

....

El Gamal

Question 27 Consider the following cryptographic scheme. Show that this scheme is homomorphic with respect to component-wise multiplications in the cipherspace? The component-wise multiplication for two vectors $\boldsymbol{x}=(x_1, x_2)$ and $\boldsymbol{y}=(y_1, y_2)$ is defined as $\boldsymbol{x} \otimes \boldsymbol{y}=(x_1 * y_1, x_2 * y_2)$.

KeyGen: For a cyclic group G of order q, sample the secret key s uniformly at random from $\mathbb{Z}_q$. For a generator g, compute the public key $p = g^s$.

Encryption: For a message $m \in G$, sample $x \in \mathbb{Z}_q$ uniformly at random and return the ciphertext $\boldsymbol{ct} = (g^x, m \cdot p^x)$.

Decryption: Form $\boldsymbol{ct} = (c_0, c_1)$. Compute the shared secret $y=c_0^s$ and compute its inverse in G. Return $m=c_1 \cdot y^{-1}$

☐0 ☐0.5 ☐1 ☐1.5 ☐2

....

....

....

Differential Privacy

Recall the definition of an ϵ-differential privacy between two neighbouring databases D, D'. Let ϵ be a positive real number. The algorithm A provides ϵ-differential privacy if, for all datasets

D, D' that differ *on a single element* and all subsets S of A

$$Pr[A(D) \in S] \leq e^{\epsilon} \times \Pr[A(D') \in S],$$

where the probability is taken over the randomness used by the differential privacy mechanism. Use this definition to answer the following questions.

Question 28 Consider a mechanism that satisfies ϵ-DP. What is the level of privacy ϵ' achieved by any group of c individuals: $Pr[M(D) \in S] \leq e^{\epsilon'} \times Pr[M(D') \in S]$, where D and D' differ at most by c individuals? Justify formally.

☐ **0** ☐ **0.5** ☐ **1** ☐ **1.5** ☐ **2**

....

....

....

Question 29 Suppose that you have a database with a single Boolean entry, and you output the following record according to coin flipping:

- If heads, you output the true value.
- If tails, you toss another coin: Output 1 if heads, and 0 otherwise.

What is the level of differential privacy in terms of ϵ that is achieved with this algorithm?

☐ **0** ☐ **0.5** ☐ **1** ☐ **1.5** ☐ **2**

....

....

....

ZKsudoku

Consider a 9x9 Sudoku grid $\mathcal{C}$ with pre-filled values as in Figure 2. The objective is to fill the grid into a solved sudoku $\mathcal{S}$ such that each subgrid[1] displays only one of each digit AND each row displays only one of each digit AND each columns displays only one of each digit. A solution can be seen on the right hand side of Figure 2. We denote by $\mathcal{S}_{i,j}$ the value in the i-th row and j-th column.

The objective is for Paul, who knows the solution to the sudoku, to prove to Vera he knows the solution without revealing it to her.

Paul and Vera engage in a sigma protocol to prove in zero-knowledge that Paul knows a solution to this particular sudoku $\mathcal{C}$.

[1] A subgrid is defined as a 3x3 grid such that the whole 9x9 grid $\mathcal{C}$ is made of 9 non-overlapping subgrids. They are delimited by bold lines on Figure 2.

	2		5		1		9	
8			2		3			6
	3			6			7	
		1				6		
5	4						1	9
		2				7		
	9			3			8	
2			8		4			7
	1		9		7		6	

a) Unsolved Sudoku $\mathcal{C}$

4	2	*6*	5	*7*	1	*3*	9	*8*
8	*5*	*7*	2	*9*	3	*1*	*4*	6
1	3	*9*	*4*	6	*8*	*2*	7	*5*
9	*7*	1	*3*	*8*	*5*	6	*2*	*4*
5	4	*3*	*7*	*2*	*6*	*8*	1	9
6	*8*	2	*1*	*4*	*9*	7	*5*	*3*
7	9	*4*	*6*	3	*2*	*5*	8	*1*
2	*6*	*5*	8	*1*	4	*9*	*3*	7
3	1	*8*	9	*5*	7	*4*	6	*2*

b) Solved Sudoku $\mathcal{S}$

Figure 2: Sudokus

First of all, Paul and Vera agree on potential splits of grid $\mathcal{S}$: into rows, columns, and subgrids. Additionally, the original public fill of the sudoku (Fig 2a) from $\mathcal{C}$ is also added to the set of potential splits. Thus, the set of splits contains the 9 columns, the 9 rows, and the 9 subgrids (3x3 grids) of the solved sudoku plus the public fill of the sudoku.

Now consider that Paul can do the following actions:
Permute. Paul creates a permutation ψ of each of the cells' value. For instance:

$$\psi : 1 \to 3 \to 9 \to 7 \to 8 \to 6 \to 4 \to 5 \to 2 \to 1$$

The result is a grid $\mathcal{S}'$ such that for all row i and column j, $\mathcal{S}'_{i,j}{=}\psi(\mathcal{S}_{i,j})$.

Mask and Commit. Paul generates a 9x9 mask grid with nonce values denoted by $mask_{i,j}$. For each cell (i,j) of an input grid $\mathcal{G}$ Paul commits to $\mathrm{Com}(\mathcal{G}_{i,j}||mask_{i,j})$, where $\mathrm{Com}(\cdot)$ is a commitment scheme. Paul outputs a 9x9 grid $\mathcal{S}''$ comprising the committed masked values: for all row i and column j:

$$\mathcal{S}''_{i,j}{=}\mathrm{Com}(\mathcal{G}_{i,j}||mask_{i,j})$$

Question 30 Using the two actions available to Paul ("permute" and "mask and commit") in the appropriate order, design a sigma-protocol for Paul to prove in zero-knowledge the knowledge of the solution to the sudoku. Each line of the response corresponding to an element of the sigma-protocol.

☐0 ☐0.5 ☐1 ☐1.5 ☐2

Question 31 What is the soundness of the above protocol: i.e., what is Vera's confidence in Paul's knowledge of the solution of the sudoku at the end of the sigma-protocol ?

☐0 ☐0.5 ☐1 ☐1.5 ☐2

....

....

....