Problem Set 1

Please submit your problem set, in PDF format, on Gradescope. *Each problem should be in a separate page.*

You are to work on this problem set in groups. For problem sets 1, 2, and 3, we will randomly assign the groups for the problem set. After problem set 3, you are to work on the following problem sets with groups of your choosing of size three or four. If you need help finding a group, try posting on Piazza. See the course website for our policy on collaboration. Each group member must independently write up and submit their own solutions.

Homework must be typeset in LaTeX and submitted electronically! Each problem answer must be provided as a separate page. Mark the top of each page with your group member names, the course number (6.5610), the problem set number and question, and the date. We have provided a template for LaTeX on the course website (see the *Psets* tab at the top of the page).

Problem 1-1. One-way functions and collision resistance

A function $f : \{0,1\}^* \to \{0,1\}^*$ is a *one-way function* (OWF) if it is computable in polynomial-time and for any polynomial-time adversary A there exists a negligible function μ such that for every $\lambda \in \mathbb{N}$,

$$\Pr\left[f(x) = f(x') \colon \begin{array}{c} x \xleftarrow{R} \{0,1\}^\lambda \\ x' \leftarrow A(f(x)) \end{array}\right] \leq \mu(\lambda).$$

In other words, given $f(x)$ it is difficult to find x' such that $f(x') = f(x)$.

A family of functions $\{f_\lambda\}_{\lambda\in\mathbb{N}}$ is said to be *collision resistant* if it is polynomial-time computable, for every $\lambda \in \mathbb{N}$, $f_\lambda : \{0,1\}^* \to \{0,1\}^\lambda$, and for all polynomial-time adversaries A there exists a negligible function μ such that for every $\lambda \in \mathbb{N}$,

$$\Pr\left[f_\lambda(x) = f_\lambda(x') \wedge x \neq x' \;:\; (x, x') \leftarrow A(1^\lambda)\right] \leq \mu(\lambda).$$

In other words, it is difficult to find distinct x, x' such that $f_\lambda(x) = f_\lambda(x')$.

For each of the following functions g determine if g is necessarily a one-way function (OWF). If so, explain in a few sentences why it is a OWF, and if not, provide an attack.

For simplicity, in what follows we define f and g for a given input length, and omit the subscript λ from the collision resistant hash functions.

(a) Let $f : \{0,1\}^n \to \{0,1\}^\lambda$ be a OWF, and let $g : \{0,1\}^n \to \{0,1\}^{2\lambda}$ where $g(x) = f(x)||0^\lambda$.

Solution:
Answer: g is a OWF.
Reasoning: Assume that given $y = f(x)||0^\lambda$, it is easy to find x such that $g(x) = y$. Then, given $y' = f(x)$, it is easy to find x' such that $g(x') = y'||0^\lambda$ and therefore $f(x') = y'$, giving us a contradiction.

(b) Let $f : \{0,1\}^\lambda \to \{0,1\}^\lambda$ be a OWF, and let $g : \{0,1\}^{\lambda/2} \to \{0,1\}^\lambda$ where $g(x) = f(x||0^{\lambda/2})$ and we assume for simplicity that λ is even.

Solution:
Answer: g is not necessarily a OWF.
Reasoning: Let us construct f such that $f(x_1||x_2) = x_1||h(x_2)$ where h is a OWF from $\{0,1\}^{\lambda/2} \to \{0,1\}^{\lambda/2}$. We know that f is a OWF using the same idea as part (a). Now we will construct an attack on g using the previously chosen OWF f. We see that $g(x) = f(x||0^{\lambda/2}) = x||h(0^{\lambda/2})$. The attacker can invert g simply by taking the first half.

(c) Let $g : \{0,1\}^* \to \{0,1\}^\lambda$ be collision resistant. Is g necessarily a OWF?

Solution:
Answer: g is a OWF.
Reasoning: Suppose a function is not one way. Then, given random x, we can find $x' \in \{0,1\}^n$ such that $g(x') = g(x)$ and $n > \lambda$. We also want to ensure that $x' \neq x$. We can simply do this by repeating the same algorithm for different random x. The reason that this works is because the input space is much larger than the output space of g. Therefore the probability of collision is very high (the probability that we fail and try again is only $1 - 1/2^{n-\lambda}$).
(optional explanation) More formally, for every value y in the output space, we can define $S(y)$ to be the number of inputs that map to y: $S(y) = |\{x \in \{0,1\}^n : g(x) = y\}|$. At each step, the probability of failure is the sum over the probability of that $g(x) = y$ for a given y times the probability of choosing $x' = x$ given that $g(x) = y$:

$$P[failure] = \sum_{y \in \{0,1\}^\lambda} P[g(x) = y] P[x' = x \text{ given } g(x) = y]$$

. The probability of getting a given y is $\frac{S(y)}{2^n}$, and the probability of choosing $x' = x$ given that $g(x) = y$ is $\frac{1}{S(y)}$. Therefore, the probability of failure is $\sum_{y \in \{0,1\}^\lambda} \frac{S(y)}{2^n} \frac{1}{S(y)} = \frac{1}{2^{n-\lambda}}$.

For each of the following functions g determine if g is necessarily a collision resistant function. If so, explain in a few sentences why it is collision resistant, and if not, provide an attack.

(d) Let $f : \{0,1\}^n \to \{0,1\}^\lambda$ be collision resistant, and let $g : \{0,1\}^{2n} \to \{0,1\}^\lambda$ where $g(x) = f(x_1||x_3||x_5||...||x_{2n-1})$.

Solution:
Answer: g is not collision resistant.
Reasoning: Any x, y such that x and y share the same odd digits will collide.

(e) Let $f : \{0,1\}^{2\lambda} \to \{0,1\}^\lambda$ be collision resistant, and let $g : \{0,1\}^{4\lambda} \to \{0,1\}^\lambda$ where $g(x_1||x_2) = f(f(x_1)||f(x_2))$.

Solution:
Answer: g is collision resistant.
Reasoning: Suppose we find distinct x, y such that $g(x) = g(y)$. We have two cases: $f(x_1) = f(y_1)$ and $f(x_2) = f(y_2)$, or at least one of the equalities do not hold. In the first case, we know that $x \neq y$, so $x_i \neq y_i$ for some i. WLOG suppose $i = 1$, therefore x_1 and y_1 can be used as a collision for f. For the second case, $(f(x_1), f(x_2))$ and $(f(y_1), f(y_2))$ can be used as a collision for f.

(f) Let $g : \{0,1\}^* \to \{0,1\}^\lambda$ be a OWF. Is g necessarily collision resistant?

Solution:
Answer: g is not necessarily collision resistant.
Reasoning: Define $g(x) = h(x[0:-1])$ for some $h : \{0,1\}^* \to \{0,1\}^\lambda$ that is a OWF. g is one-way because given y, if we find x' such that $g(x') = y$, then $h(x'[0:-1]) = y$. However, g is not collision resistant because for any x, $g(x||0) = g(x||1)$.

Problem 1-2. Short integer solutions

We will need the following definition:

> **Definition** (Short integer solutions (SIS))**.** The short-integer-solutions problem is parameterized by positive integers n, m, q, and B. For a random matrix $\mathbf{A} \xleftarrow{\text{R}} \mathbb{Z}_q^{n \times m}$, the problem is to find a nonzero vector $\mathbf{e} \in \mathbb{Z}^m$ such that:

(1) $\mathbf{A} \cdot \mathbf{e} = \mathbf{0} \in \mathbb{Z}_q^n$ and
(2) $\|\mathbf{e}\|_\infty \leq B$, where $\|\cdot\|_\infty$ denotes the L_∞-norm where $\|\mathbf{x}\|_\infty = \max |x_i|$

To be fully formal, we can treat n as the security parameter and then let m, q, and B be functions of n. (Often, we will take all of these parameters to be polynomials in n.) An example parameter setting might be $n = 1024$, $q = 2^{32}$, $m = 4n \log q$, and $B = 1$.

Then we have:

Definition (SIS assumption)**.** The SIS assumption on parameters (n, m, q, B) is that for all p.p.t. adversaries $\mathcal{A}$, there is a negligible function $\mu(\cdot)$ such that

$$\Pr\left[\mathbf{A} \cdot \mathbf{e} = \mathbf{0} \wedge \|\mathbf{e}\|_\infty \leq B \;:\; \begin{array}{c} \mathbf{A} \xleftarrow{\mathrm{R}} \mathbb{Z}_q^{n \times m} \\ \mathbf{e} \leftarrow \mathcal{A}(\mathbf{A}) \end{array}\right] \leq \mu(n).$$

(a) For the SIS assumption to hold, the parameters (n, m, q, B) need to satisfy certain conditions. We will list a few insecure settings of the SIS parameters. For each, explain why we do not instantiate the SIS problem with this parameter setting:

(a) $(n, 2n, 2, 2)$

Solution: Easy to solve with Gaussian elimination. Could also set $\mathbf{e}$ to be the zero vector using 2 in place of 0.

(b) $(n, n/1000, q, 1)$, for a prime $q \approx n$ (*Hint:* Count the number of possible inputs and outputs.)

Solution: There is often no solution.

(c) $(n, 10n \log n, q, 2)$ for a prime $q \approx n$, except rather than sampling $\mathbf{A}$ at random from $\mathbb{Z}_q^{n \times m}$, we sample a random matrix $\mathbf{A}$ that has at most one non-zero element in each row and at most one non-zero element in each column.

Solution: It is easy to find a solution by constructing the solution $\mathbf{e}$ one entry at a time.

(b) For SIS parameters (m, n, q, B) and a random matrix $\mathbf{A} \xleftarrow{\mathrm{R}} \mathbb{Z}_q^{n \times m}$, let $H_{\mathbf{A}} \colon \{0,1\}^m \to \mathbb{Z}_q^n$ be a hash function, defined as $H_{\mathbf{A}}(\mathbf{e}) := \mathbf{A} \cdot \mathbf{e}$. Explain why $H_{\mathbf{A}}$ is collision resistant under the SIS assumption with parameters $(m, n, q, 2B)$. In other words, given an efficient algorithm $\mathcal{A}$ that finds a collision in $H_{\mathbf{A}}$, produce an efficient algorithm that breaks the SIS assumption with the given parameters.

Solution: Say that $\mathbf{e}$ and $\mathbf{e}'$ collide under $H_{\mathbf{A}}$. Then $\mathbf{A} \cdot \mathbf{e} = \mathbf{A} \cdot \mathbf{e}'$ and therefore $\mathbf{A} \cdot (\mathbf{e} - \mathbf{e}') = \mathbf{0}$. Now $\mathbf{e} - \mathbf{e}'$ has L_∞ norm at most $2B$ and thus is a solution to the SIS problem with parameters $(m, n, q, 2B)$.

(c) State two reasons (in at most one sentence each) why we might not use $H_{\mathbf{A}}$ as a collision-resistant hash function in practice. Again, think of the parameters necessary to achieve 128-bit security as being something like $n = 2^{10}$, $q = 2^{32}$, $m = 2^{17}$, and $B = 1$.

Solution: Computation time is at least quadratic in the security parameter.
Describing the matrix $\mathbf{A}$ requires a large number of bits. So the code of the hash function itself will be large.

Problem 1-3. Finite fields and polynomials In this problem we consider polynomials over finite fields and finite rings. Specifically, let $\mathbf{Z}_n$ be the ring of elements $\{0, 1, \ldots, n-1\}$ where addition and multiplication are done modulo n. When n is a prime this is a field (where every non-zero element has a multiplicative inverse) and when n is not a prime it is a ring (where not all elements have a multiplicative inverse). In this problem we consider degree d polynomials $f : \mathbf{Z}_n \to \mathbf{Z}_n$. Such polynomials can be represented as $f(x) = \sum_{i=0}^{d} a_i x^i$ where $a_0, \ldots, a_d \in \mathbf{Z}_n$ and where addition and multiplication are done modulo n.

(a) Argue that a non-zero degree-d polynomial modulo a prime p has $\leq d$ roots. (Hint: There are multiple ways to do this. One way is by induction. Another way uses the fact that the Vandermonde matrix over a field has full rank.)

Solution: Proof by induction on d. For $d = 1$ and for any $f = a_1x + a_0$ a root is x such that $a_1x + a_0 = 0$. Since $a_1 \neq 0$ it has an inverse which implies that $x = -a_0 \cdot a_1^{-1}$ and thus is unique. Suppose this is true for $d-1$ we prove that it is true for d. Given any $f = \sum_{i=0}^{d} a_i x^i$ let x_1 be one of its roots. Then f can be written as $f(x) = (x - x_1) \cdot g(x)$ where g is of degree $d - 1$. By our induction assumption g has at most $d - 1$ roots. Thus, in total f has at most d roots.

(b) Argue that for every prime p, every distinct $x_1, \ldots, x_{d+1} \in \mathbf{Z}_p$ and every $y_1, \ldots, y_{d+1} \in \mathbf{Z}_p$, there exists a unique degree-d polynomial $f : \mathbf{Z}_p \to \mathbf{Z}_p$ such that $f(x_i) = y_i$ for every $i \in \{1, \ldots, d+1\}$. (Hint: You can assume that the Vanderonde matrix has full rank.)

Solution: Uniqueness follows directly from part (a), since if there exist two degree-d polynomials f_1 and f_2 that agree on $x_1, \ldots, x_{d+1}$. This implies that $g = f_1 - f_2$ is a non-zero polynomial of degree $\leq d$ that has $d+1$ roots, contradicting part (a). We next argue that there exist $a_0, \ldots, a_d \in \mathbf{Z}_p$ such that for every $j \in [d+1]$ it holds that $\sum a_i x_j^i = y_j$. These can be seen as $d+1$ linear equations in $d+1$ variables (the variables being $a_0, a_1, \ldots, a_d$). It has a unique solution.

(c) Give an example of a (non-prime) n and a degree-d polynomial $f : \mathbf{Z}_n \to \mathbf{Z}_n$ that has more than d roots.

Solution: Let $n = 4$ and $f(x) = 2x$. Both $x = 0$ and $x = 2$ are roots.

Problem 1-4. Breaking AES without S-box

On piazza, you can find a zip file `pset1.zip` that contains the file needed for this problem.

`gen.py` contains an AES encryption implementation, except the substitution operations are omitted. Using this wrong implementation, it encrypts the secret message (`secret.txt`) and 150 random blocks. Your goal is to recover the secret.

- `secret.txt` is the secret you want to recover and is not contained in the zip file.
- `gen.py` reads from `secret.txt` and generates `ciphertext.txt` and `data.txt`. You don't have to read the details of AES encryption as it's meant to be a correct implementation (except for the substitution); it should be enough to check the code in the main function.
- `ciphertext.txt` is the encrypted secret. It contains non-ASCII characters, so don't be surprised if it looks garbled.
- `data.txt` has 150 lines, each containing a block and its encryption in hex.
- `hint.pdf` provides hints and guidance, but please try it without checking the hints first!

Please submit the secret message and the code on Gradescope.

Solution: In block ciphers, S-box is typically used to obscure the relationship between the key and the ciphertext, thus ensuring Shannon's property of confusion. Mathematically, an S-box is a nonlinear vectorial Boolean function. (https://en.wikipedia.org/wiki/S-box)