Theoretical Foundations of Cryptography
 Georgia Tech, Spring 2011

Homework 3

Instructor: Chris Peikert
Student: YOUR NAME HERE

This homework is due by the **start of class on February 24** via the course page on T-Square. Start early!

Instructions. Solutions must be typeset in LaTeX (a template for this homework is available on the course web page). Your work will be graded on *correctness*, *clarity*, and *conciseness*. You should only submit work that you believe to be correct; if you cannot solve a problem completely, you will get significantly more partial credit if you clearly identify the gap(s) in your solution. It is good practice to start any long solution with an informal (but accurate) "proof summary" that describes the main idea.

You may collaborate with others on this problem set and consult external sources. However, you must ***write your own solutions*** and ***list your collaborators/sources*** for each problem.

1. (10 points) Recall that multi-message *non-adaptive* security for a symmetric-key encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ says that for any $q = \mathrm{poly}(n)$ and any tuples $(m_1, \ldots, m_q), (m'_1, \ldots, m'_q) \in \mathcal{M}^q$, it should be the case that

$$\{k \leftarrow \mathsf{Gen} : (\mathsf{Enc}_k(m_1), \ldots, \mathsf{Enc}_k(m_q))\} \overset{c}{\approx} \left\{k \leftarrow \mathsf{Gen} : (\mathsf{Enc}_k(m'_1), \ldots, \mathsf{Enc}_k(m'_q))\right\}.$$

In contrast, *adaptive* security says that the following two oracles are indistinguishable:

$$\left\langle k \leftarrow \mathsf{Gen} : \mathsf{Enc}^0_k(\cdot, \cdot)\right\rangle \overset{c}{\approx} \left\langle k \leftarrow \mathsf{Gen} : \mathsf{Enc}^1_k(\cdot, \cdot)\right\rangle,$$

where $\mathsf{Enc}^b_k(m_0, m_1)$ outputs $\mathsf{Enc}_k(m_b)$. Give a separation between these two definitions, i.e., construct a (possibly contrived) scheme and prove it secure according to the former definition (under some standard assumption), while showing that it is definitely insecure according to the latter definition.

2. In a network where all communication is via broadcast (e.g., wi-fi), we might want communication to be *anonymous*. That is, the adversary should not be able to learn, given an (encrypted) message, anything about who the sender and intended receiver are.

   (a) (5 points) Give a simple but formal definition of *anonymity* under adaptive chosen-plaintext attack for a symmetric-key encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$. The syntax of your definition should capture *only* the intuition that it is hard to distinguish which key was used to produce the queried ciphertexts.

   (b) (5 points) Does your anonymity definition imply our standard security notion of IND-CPA (indistinguishability under adaptive chosen-plaintext attack)? If so, prove it; otherwise, give the simplest counterexample you can find, *and* provide a definition that captures both anonymity and IND-CPA. (In constructing your counterexample, you may rely on any reasonable cryptographic assumption).

   (c) (5 points) Does the PRF-based encryption scheme from class satisfy your definition of anonymity? If so, prove it; otherwise, describe a PRF family for which the encryption scheme is not anonymous. (As usual, your counterexample may use any reasonable cryptographic assumption.)

3. (15 points) *(The power of Decision Diffie-Hellman.)* In class we saw that the DDH assumption can be used for public-key encryption; here you will show that it is very useful for *symmetric* primitives too.

   For a cyclic group $G = \langle g \rangle$ of *prime* order q, the DDH assumption says that

$$(g, g^a, g^b, g^{ab}) \overset{c}{\approx} (g, g^a, g^b, g^c),$$

where $a, b, c \leftarrow \mathbb{Z}_q$ are uniformly random and independent. By grouping the elements appropriately, we can view this assumption in matrix form:

$$g^{\begin{pmatrix} 1 & a \\ 1 \cdot b & a \cdot b \end{pmatrix}} \overset{c}{\approx} g^{\begin{pmatrix} 1 & a \\ b & c \end{pmatrix}},$$

where g^M (for a matrix M over $\mathbb{Z}_q$) is the matrix over G obtained by raising g to each entry of M. Observe that in the left-hand matrix, the two rows are linearly dependent (over $\mathbb{Z}_q$), while in the right-hand matrix they are very likely not to be.

(a) (7 points) Prove that the DDH assumption implies that, for any positive integer $w = \mathrm{poly}(n)$,

$$g^{\begin{pmatrix} a_1 & a_2 & \cdots & a_w \\ a_1 \cdot b & a_2 \cdot b & \cdots & a_w \cdot b \end{pmatrix}} \overset{c}{\approx} g^{\begin{pmatrix} a_1 & a_2 & \cdots & a_w \\ c_1 & c_2 & \cdots & c_w \end{pmatrix}},$$

where $a_i, b, c_i \leftarrow \mathbb{Z}_q$ are all uniformly random and independent.

(b) (4 points) Using the previous part, prove that the DDH assumption implies that, for any positive integers $w, h = \mathrm{poly}(n)$,

$$g^{\left(a_i \cdot b_j\right)_{i \in [h], j \in [w]}} \overset{c}{\approx} g^{\left(c_{i,j}\right)_{i \in [h], j \in [w]}},$$

where $a_i, b_j, c_{i,j} \leftarrow \mathbb{Z}_q$ are all uniformly random and independent. Note that the left-hand matrix (in the exponent) has rank 1, while the right-hand matrix is very likely to be full-rank.

(c) (4 points) Conclude that under the DDH assumption, there is a PRG family expanding about $2n \lg q$ bits to about $n^2 \lg q$ bits. (The output need not literally be made up of bits, though.) For the same input and output lengths, why might we prefer this PRG to the one from class based on a OWP?

(d) (0 points) *(Challenge question.)* Generalize the above to design a pseudorandom *function* based on DDH. (*Hint*: extend to $2 \times 2 \times \cdots \times 2$ matrices.)