MTAT.07.003 Cryptology II
Spring 2012 / Exercise session ?? / Example Solution

Exercise (Security of hash ElGamal cryptosystem). *The message space of the ElGamal cryptosystem is a DDH group* $\mathbb{G}$. *The latter is rather limiting, since normally one needs to encrypt n-bit messages and not the group elements. The hash ElGamal cryptosystem for q-element group* $\mathbb{G} = \langle g \rangle$ *is defined as follows:*

$$\begin{array}{lll} \mathsf{Gen} & \mathsf{Enc}_{\mathsf{pk}}(m) & \mathsf{Dec}_{\mathsf{sk}}(c_1, c_2) \\ \left[\begin{array}{l} x \xleftarrow{u} \mathbb{Z}_q \\ y \leftarrow g^x \\ \textbf{\textit{return}}\ (x, y) \end{array}\right. & \left[\begin{array}{l} k \xleftarrow{u} \mathbb{Z}_q \\ c_1 \leftarrow g^r \\ c_2 \leftarrow m \oplus h(y^r) \\ \textbf{\textit{return}}\ (c_1, c_2) \end{array}\right. & \left[\begin{array}{l} m \leftarrow c_2 \oplus h(c_1^x) \\ \textbf{\textit{return}}\ m\ . \end{array}\right. \end{array}$$

where the secret key is x and the public key is y, and $h : \mathbb{G} \to \{0, 1\}^n$ *is a almost regular hash function. That is, the distribution* $h(y)$ *for* $y \xleftarrow{u} \mathbb{G}$ *is statistically* ε_2*-close to the uniform distribution over* $\{0, 1\}^n$. *Prove that the simplified ElGamal cryptosystem is also IND-CPA secure and give the corresponding security bounds.*

SOLUTION. Let us consider the IND-CPA games for the simplified ElGamal cryptosystem. For brevity, let q denote the size of the group $\mathbb{G}$.

$$\begin{array}{ll} \mathcal{G}_0^{\mathcal{A}} & \mathcal{G}_1^{\mathcal{A}} \\ \left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ k \leftarrow \mathbb{Z}_q \\ c \leftarrow (g^k, h(g^{xk}) \oplus m_0) \\ \textbf{return}\ \mathcal{A}(c) \end{array}\right. & \left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ k \leftarrow \mathbb{Z}_q \\ c \leftarrow (g^k, h(g^{xk}) \oplus m_1) \\ \textbf{return}\ \mathcal{A}(c) \end{array}\right. \end{array}$$

Using the fact that $\mathbb{G}$ is (t, ε_1)-secure DDH group, we get another pair of games such that the distance between $\mathcal{G}_0$ and $\mathcal{G}_2$ and $\mathcal{G}_1$ and $\mathcal{G}_3$ is ε_1.

$$\begin{array}{ll} \mathcal{G}_2^{\mathcal{A}} & \mathcal{G}_3^{\mathcal{A}} \\ \left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ k \leftarrow \mathbb{Z}_q \\ \ell \leftarrow \mathbb{Z}_q \\ c \leftarrow (g^k, h(g^{\ell}) \oplus m_0) \\ \textbf{return}\ \mathcal{A}(c) \end{array}\right. & \left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ k \leftarrow \mathbb{Z}_q \\ \ell \leftarrow \mathbb{Z}_q \\ c \leftarrow (g^k, h(g^{\ell}) \oplus m_1) \\ \textbf{return}\ \mathcal{A}(c) \end{array}\right. \end{array}$$

Indeed, note that the corresponding game pairs differ in a single line—we have replaced group element g^{xk} by a random group element g^{ℓ}. As a result, if there is a significant change in the success of adversary $\mathcal{A}$, we can easily construct an adversary $\mathcal{B}$ against DDH problem defined by the following game pair:

$$\begin{array}{ll} \mathcal{Q}_0^{\mathcal{B}} & \mathcal{Q}_1^{\mathcal{B}} \\ \left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ y \leftarrow \mathbb{Z}_q \\ z \leftarrow xy \\ \textbf{return}\ \mathcal{B}(g, g^x, g^y, g^z) \end{array}\right. & \left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ y \leftarrow \mathbb{Z}_q \\ z \leftarrow \mathbb{Z}_q \\ \textbf{return}\ \mathcal{B}(g, g^x, g^y, g^z)\ . \end{array}\right. \end{array}$$

We know that for any t-time adversary $\mathcal{B}$ the advantage $\mathsf{Adv}^{\mathsf{ddh}}_{\mathbb{G}}(\mathcal{B}) \leq \varepsilon_1$. However if we consider the following adversaries.

$$\begin{array}{ll}
\mathcal{B}_1(g, g^x, g^y, g^z) & \mathcal{B}_2(g, g^x, g^y, g^z) \\
\left[\begin{array}{l} \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ c \leftarrow (g^y, h(g^z) \oplus m_0) \\ \mathbf{return}\ \mathcal{A}(c) \end{array}\right. &
\left[\begin{array}{l} \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ c \leftarrow (g^y, h(g^z) \oplus m_1) \\ \mathbf{return}\ \mathcal{A}(c) \end{array}\right.
\end{array}$$

By inserting $\mathcal{B}_1$ to DDH game $\mathcal{Q}_0$ we get a game that is identical to $\mathcal{G}_0^{\mathcal{A}}$ and by inserting $\mathcal{B}_1$ to $\mathcal{Q}_1$ we get a game that is identical to $\mathcal{G}_2^{\mathcal{A}}$. Similarly, $\mathcal{Q}_0^{\mathcal{B}_2} \equiv \mathcal{G}_1^{\mathcal{A}}$ and $\mathcal{Q}_1^{\mathcal{B}_2} \equiv \mathcal{G}_3^{\mathcal{A}}$. Consequently, we can conclude that

$$\begin{aligned}
\mathsf{Adv}^{\mathsf{ddh}}_{\mathbb{G}}(\mathcal{B}_1) &= \left|\Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_2^{\mathcal{A}} = 1\right]\right| \\
\mathsf{Adv}^{\mathsf{ddh}}_{\mathbb{G}}(\mathcal{B}_2) &= \left|\Pr\left[\mathcal{G}_1^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_3^{\mathcal{A}} = 1\right]\right| \enspace .
\end{aligned}$$

Since the running times of $\mathcal{B}_1$ and $\mathcal{B}_2$ are comparable to the running time of $\mathcal{A}$, we we have proved that for any t-time $\mathcal{A}$:

$$\begin{aligned}
\left|\Pr\left[\mathcal{G}_0^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_2^{\mathcal{A}} = 1\right]\right| &\leq \varepsilon_1 \\
\left|\Pr\left[\mathcal{G}_1^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_3^{\mathcal{A}} = 1\right]\right| &\leq \varepsilon_1 \enspace .
\end{aligned}$$

As g^ℓ is a uniform element of $\mathbb{G}$, we can use almost regularity of h. More precisely, we can define a pair of games $\mathcal{G}_4$ and $\mathcal{G}_5$ such that the corresponding statistical distance from $\mathcal{G}_2$ and $\mathcal{G}_3$ is below ε_2:

$$\begin{array}{lll}
\mathcal{G}_4^{\mathcal{A}} & \mathcal{G}_6^{\mathcal{A}} & \mathcal{G}_5^{\mathcal{A}} \\
\left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ k \leftarrow \mathbb{Z}_q \\ r \leftarrow \{0,1\}^n \\ c \leftarrow (g^k, r \oplus m_0) \\ \mathbf{return}\ \mathcal{A}(c) \end{array}\right. &
\left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ k \leftarrow \mathbb{Z}_q \\ r \leftarrow \{0,1\}^n \\ c \leftarrow (g^k, r) \\ \mathbf{return}\ \mathcal{A}(c) \end{array}\right. &
\left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ k \leftarrow \mathbb{Z}_q \\ r \leftarrow \{0,1\}^n \\ c \leftarrow (g^k, r \oplus m_1) \\ \mathbf{return}\ \mathcal{A}(c) \enspace . \end{array}\right.
\end{array}$$

For the formal reasoning, note that by the properties of h we know that $\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{X}_0,\mathcal{X}_1}(\mathcal{B}) \leq \varepsilon_2$ for any imaginable adversary $\mathcal{B}$, where the samples of $\mathcal{X}_0$ are generated $h(g)$ for $g \leftarrow \mathbb{G}$ and $\mathcal{X}_1$ is uniform distribution over $\{0,1\}^n$. Hence, we have to construct an adversary $\mathcal{B}$ for the distinguishing games

$$\begin{array}{ll}
\mathcal{Q}_0^{\mathcal{B}} & \mathcal{Q}_1^{\mathcal{B}} \\
\left[\begin{array}{l} x \leftarrow \{0,1\}^n \\ \mathbf{return}\ \mathcal{B}(x) \end{array}\right. &
\left[\begin{array}{l} y \leftarrow \mathbb{Z}_q \\ x \leftarrow h(g^y) \\ \mathbf{return}\ \mathcal{B}(x) \end{array}\right.
\end{array}$$

in order to formally prove distance bounds for the game pairs $(\mathcal{G}_2, \mathcal{G}_4)$ and $(\mathcal{G}_3, \mathcal{G}_5)$. Now note that the adversary constructions

$$\begin{array}{ll}
\mathcal{B}_1(r) & \mathcal{B}_2(r) \\
\left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ k \leftarrow \mathbb{Z}_q \\ c \leftarrow (g^k, r \oplus m_0) \\ \mathbf{return}\ \mathcal{A}(c) \end{array}\right. &
\left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ k \leftarrow \mathbb{Z}_q \\ c \leftarrow (g^k, r \oplus m_1) \\ \mathbf{return}\ \mathcal{A}(c) \end{array}\right.
\end{array}$$

assure that

$$\begin{aligned} \mathcal{Q}_0^{\mathcal{B}_1} &\equiv \mathcal{G}_4^{\mathcal{A}} & \qquad \mathcal{Q}_0^{\mathcal{B}_2} &\equiv \mathcal{G}_5^{\mathcal{A}} \\ \mathcal{Q}_1^{\mathcal{B}_1} &\equiv \mathcal{G}_2^{\mathcal{A}} & \qquad \mathcal{Q}_1^{\mathcal{B}_2} &\equiv \mathcal{G}_3^{\mathcal{A}} \end{aligned}$$

and thus

$$\begin{aligned} \left|\Pr\left[\mathcal{G}_2^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_4^{\mathcal{A}} = 1\right]\right| &\le \varepsilon_2 \\ \left|\Pr\left[\mathcal{G}_3^{\mathcal{A}} = 1\right] - \Pr\left[\mathcal{G}_5^{\mathcal{A}} = 1\right]\right| &\le \varepsilon_2 \ . \end{aligned}$$

To complete the proof, note that two games $\mathcal{G}_4$ and $\mathcal{G}_5$ are equivalent to game $\mathcal{G}_6$ because in both cases $\mathcal{A}$ gets a random element g^k, r.Combining the results with the help of triangle inequality, we get that distance of $\mathcal{G}_0$ and $\mathcal{G}_1$ is at most $2\varepsilon_1 + 2\varepsilon_2$. As the allowed running-time for $\mathcal{A}$ is bounded by the (t, ε_1)-secure Decisional Diffie-Hellman group property used in the reduction, so the previously defined advantage holds for any t-time adversary $\mathcal{A}$. Hence, the simplified ElGamal is $(t, 2\varepsilon_1 + 2\varepsilon_2)$ IND-CPA secure.

Computational uniformity. The function h does not have to be almost uniform. The reduction constructions $\mathcal{B}_1$ and $\mathcal{B}_2$ for the game pairs $(\mathcal{G}_2, \mathcal{G}_4)$ and $(\mathcal{G}_3, \mathcal{G}_5)$ are very efficient – the running times of $\mathcal{B}_1$ and $\mathcal{B}_2$ are comparable to the running time of $\mathcal{A}$. Hence, cryptographic assumptions on h can be relaxed. It is sufficient that $\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{X}_0,\mathcal{X}_1}(\mathcal{B}) \le \varepsilon_2$ for all t-time adversaries $\mathcal{B}$.

Direct constructive proof. The other way to think about the problem is that if $\mathcal{A}$ is very good against IND-CPA games, $\mathcal{A}$ must be also good for the game

$$\begin{aligned} &\mathcal{G}^{\mathcal{A}} \\ &\left[\begin{array}{l} x \leftarrow \mathbb{Z}_q \\ \mathsf{pk} \leftarrow g^x \\ m_0, m_1 \leftarrow \mathcal{A}(\mathsf{pk}) \\ i \leftarrow \{0,1\} \\ k \leftarrow \mathbb{Z}_q \\ c \leftarrow (g^k, h(g^{xk}) \oplus m_i) \\ guess \leftarrow \mathcal{A}(c) \\ \textbf{return } [guess \stackrel{?}{=} i] \ . \end{array}\right. \end{aligned}$$

As a result, we can use $\mathcal{A}$ directly for distinguishing DDH games by defining the following adversary

$$\begin{aligned} &\mathcal{B}(g, g^x, g^k, y) \\ &\left[\begin{array}{l} m_0, m_1 \leftarrow \mathcal{A}(g^x) \\ i \leftarrow \{0,1\} \\ c \leftarrow (g^k, h(y) \oplus m_i) \\ guess \leftarrow \mathcal{A}(c) \\ \textbf{return } [guess \stackrel{?}{=} i] \end{array}\right. \end{aligned}$$

If $\mathcal{B}$ plays against DDH game $\mathcal{Q}_0^{\mathcal{B}}$ then $y = g^{kx}$ and therefore $\mathcal{A}$ will play the game $\mathcal{G}$. By using the well-known equivalence between the IND-CPA games and the guessing game $\mathcal{G}$, we get

$$\begin{aligned} \Pr\left[\mathcal{Q}_0^{\mathcal{B}} = 1\right] &= \frac{1}{2} \cdot \Pr\left[c = (g^k, h(g^{xk}) \oplus m_0) : \mathcal{A}(c) = 0\right] \\ &\quad + \frac{1}{2} \cdot \Pr\left[c = (g^k, h(g^{xk}) \oplus m_1) : \mathcal{A}(c) = 1\right] \\ &= \frac{1}{2} \pm \frac{1}{2} \cdot \mathsf{Adv}^{\mathsf{ind\text{-}cpa}}(\mathcal{A}) \ . \end{aligned}$$

In the game $\mathcal{Q}_1$, however, $y = g^\ell$ is a random group element and we obtain

$$\begin{aligned} \Pr\left[\mathcal{Q}_1^{\mathcal{B}} = 1\right] &= \frac{1}{2} \cdot \Pr\left[c = (g^k, h(y) \oplus m_0) : \mathcal{A}(c) = 0\right] \\ &\quad + \frac{1}{2} \cdot \Pr\left[c = (g^k, h(y) \oplus m_1) : \mathcal{A}(c) = 1\right] \\ &= \frac{1}{2} \pm \frac{1}{2} \cdot \mathsf{Adv}^{\mathsf{ind}}_{\mathcal{Y}_0,\mathcal{Y}_1}(\mathcal{A}) \end{aligned}$$

where $\mathcal{Y}_0$ and $\mathcal{Y}_1$ are distributions of $h(y) \oplus m_0$ and $h(y) \oplus m_1$, respectively.

Since y is a uniformly chosen group element then the distributions $h(y) \oplus m_0$ and $h(y) \oplus m_1$ are at most ε_2 apart from uniform distribution over $\{0,1\}^n$. By the triangle inequality we can conclude

$$\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{Y}_0,\mathcal{Y}_1}(\mathcal{A}) \leq \varepsilon_2 \ .$$

By combining both estimates, we get

$$\mathsf{Adv}^{\mathsf{ddh}}_{\mathbb{G}}(\mathcal{B}) \geq \frac{1}{2} \cdot \mathsf{Adv}^{\mathsf{ind\text{-}cpa}}(\mathcal{A}) - \mathsf{Adv}^{\mathsf{ind}}_{\mathcal{Y}_0,\mathcal{Y}_1}(\mathcal{A}) \geq \frac{1}{2} \cdot \mathsf{Adv}^{\mathsf{ind\text{-}cpa}}(\mathcal{A}) - \varepsilon_2 \ .$$

Hence, if $\mathsf{Adv}^{\mathsf{ind\text{-}cpa}}(\mathcal{A}) > 2\varepsilon_1 + 2\varepsilon_2$ then the advantage $\mathsf{Adv}^{\mathsf{ddh}}_{\mathbb{G}}(\mathcal{B}) > \varepsilon_1$, which contradicts the DDH assumption, as running times of $\mathcal{A}$ and $\mathcal{B}$ are comparable.