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

Exercise (Success amplification by majority voting). *Let $\mathbb{G}$ be a finite q-element group such that all elements $y \in \mathbb{G}$ can be expressed as powers of $g \in \mathbb{G}$. Let $\mathcal{A}$ be an algorithm for finding the most significant bit of discrete logarithm such that* $\Pr[\mathcal{A}(y)$ *guesses correctly*$] \geq \varepsilon > \frac{1}{2}$ *for any $y \in \mathbb{G}$. Construct an algorithm that fails with probability 2^{-n}. Show that it is possible to give a construction withe the running-time that is linear in n and quadratic in $1/(\varepsilon - \frac{1}{2})$.*

Solution. Simple amplification. According to the assumptions the probability that $\mathcal{A}(y)$ returns correctly the most significant bit is at least $\varepsilon > \frac{1}{2}$ for all $y \in \mathbb{G}$. This assumption automatically excludes probability that $\mathcal{A}$ is a deterministic algorithm. Indeed, if $\mathcal{A}$ is deterministic then for any y it either outputs a correct answer or not. As the probability of outputting the correct answer is nonzero for all $y \in \mathbb{G}$, the deterministic $\mathcal{A}$ must output the correct output for all $y \in \mathbb{G}$ and there is nothing for us to do further. If $\mathcal{A}$ is a randomised algorithm, then depending on the randomness we get sometimes correct and sometimes incorrect answers for a fixed input y. By the assumption the fraction of correct answers is at least ε. In particular, not that if we run $\mathcal{A}(y)$ twice with freshly chosen randomness we get two independent samples from the seth of all answers. Therefore, we can define the amplification algorithm as follows:

$$\begin{array}{l}\mathcal{B}^{\mathcal{A}}(m,y)\\ \left[\begin{array}{l}\text{For } i \in \{1\ldots m\} \text{ do}\\ \left[\, x_i \leftarrow \mathcal{A}(y)\right.\\ s \leftarrow x_1 + \cdots + x_m\\ \textbf{return } [2\cdot s > m]\ .\end{array}\right.\end{array}$$

Now recall the Hoeffding bound. Let $X_1, \ldots, X_m$ be independent samples form a fixed zero-one distribution such that the probability of one is α. Then the probability that the sum of these individual samples $S = X_1 + \cdots + X_m$ is significantly less than mathematical expectation $\mathbf{E}(S)$ is negligible:

$$\Pr[\mathbf{E}(S) - S \leq m \cdot \delta] \leq \exp\left(-2m\delta^2\right) .$$

For the analysis let us consider the case, when the the correct answer is one. Then by our assumption the probability that $\mathcal{A}(y)$ returns one is at least ε. On the same time $\mathcal{B}$ returns one only if the majority of x_i-s are ones. That is we can express the failure probability as follows:

$$\begin{aligned}\Pr[x_1 + \cdots + x_m \leq m/2] &= \Pr[m\varepsilon - (x_1 + \cdots + x_m) \leq m\varepsilon - m/2]\\ &\leq \Pr[\mathbf{E}(x_1 + \cdots + x_m) - (x_1 + \cdots + x_m) \leq m(\varepsilon - 1/2)] \ .\end{aligned}$$

As the right-hand side of the inequality corresponds to the left-hand side of the Hoeffding bound, we get

$$\Pr[x_1 + \cdots + x_m \leq m/2] \leq \exp\left(-2m(\varepsilon - 1/2)^2\right)$$

Thus, we can guarantee that the failure probability is below 2^{-n} if

$$\exp\left(-2m(\varepsilon - 1/2)^2\right) \leq 2^{-n} \quad \iff \quad n \ln 2 \leq 2m(\varepsilon - 1/2)^2 \ .$$

The latter provides a lower bound for required samples:

$$m \geq \frac{n \ln 2}{2(\varepsilon - 1/2)^2} \ ,$$

which is indeed linear in n and quadratic in $1/(\varepsilon - \frac{1}{2})$. The analysis of the case where the correct answer is zero is symmetrical — again the decision bound $m/2$ is quite far from the expected number of ones.

Construction of the discrete logarithm solver. Recall that it was possible to reconstruct the full discrete logarithm if we had a perfect solver $\mathcal{B}_\circ$ for the most significant bit. Let us quickly recall the corresponding construction $\mathcal{C}$ under the assumption that the size of $\mathbb{G}$ is below 2^k. Let $y = g^x$ where $x = x_k \dots x_0$ in binary. Let $\mathsf{msb}(x) = x_k$ denote the most significant bit of x. Then clearly

$$y_1 = g^{x_{k-1} \dots x_0 0} = y \cdot g^{\mathsf{msb}(x)}$$

and we can use the most significant bit solver $\mathcal{B}_\circ$ for y_1 to recover x_{k-1}. By repeating this procedure, we can recover all bits of x by making k calls to $\mathcal{B}_\circ$:

$$\begin{array}{l}\mathcal{C}^{\mathcal{B}_\circ}(y) \\ \left[\begin{array}{l} \text{For } i = k, \dots, 0 \text{ do} \\ \left[\begin{array}{l} x_i \leftarrow \mathcal{A}_1(y) \\ y \leftarrow y^2 g^{-2x_i}) \end{array}\right. \\ \textbf{return } x_k \dots x_0 \ . \end{array}\right.\end{array}$$

If the solver $\mathcal{B}$ for the most significant bit is guaranteed to succeed with probability at least δ for any $y \in \mathbb{G}$, then it reconstructs the correct answer with the probability at least δ^k. To get a bigger success probability, we can use standard discrete logarithm amplification technique for $\mathcal{C}$. Due to the quasi-linearity of this amplification scheme, ℓ repetitions of $\mathcal{C}$ increases the success probability approximately ℓ times.

This leads us to an interesting tradeoff issue. Given an initial solver $\mathcal{A}$ for the most significant bit, we can first amplify its success by constructing the majority vote amplifier $\mathcal{B}$ with m-fold repetition and then doing an additional amplification by running ℓ times the discrete logarithm solver $\mathcal{C}$. As a result, different choices of m and ℓ can lead to the same success probability. Let us analyse the situation in more detail to determine the optimal ratio between parameters. First, note that for fixed ε and m the success probability

$$\delta \geq 1 - \exp\left(-2m(\varepsilon - 1/2)^2\right)$$

and thus the overall failure probability after ℓ reruns of $\mathcal{C}$ is not larger than

$$\Pr\left[\mathsf{Failure}\right] = \left(1 - \left(1 - \exp\left(-2m(\varepsilon - 1/2)^2\right)\right)^k\right)^\ell \approx \left(k \cdot \exp\left(-2m(\varepsilon - 1/2)^2\right)\right)^\ell \ ,$$

which itself implies

$$\log \Pr\left[\mathsf{Failure}\right] \approx \ell \cdot \log k - 2\ell \cdot m(\varepsilon - 1/2)^2 \ .$$

By looking to the equation, we see that the second term remains constant as long as $\ell \cdot m$ remains constant and the first terms increases when we increase ℓ. Consequently, an approximately optimal solution is to choose $\ell = 1$ and choose m large enough to get the desired failure probability.

On the random self-reducibility of the most significant bit. All these reductions so far assume that the success probability $\mathcal{A}$ is uniformly large for any $y \in \mathbb{G}$. In practice, we might encounter an algorithm, for which the probability of correct answer is $\varepsilon > \frac{1}{2}$ only if y is chosen uniformly form $\mathbb{G}$. Hence, we might ask is it possible to convert a particular most significant bit instance to a random most significant bit instance.

This seems to be a difficult task for the following reason. Let $x = x_k \dots x_0$ and $\overline{x} = \overline{x}_k \dots \overline{x}_0$. Then the standard rerandomisation procedure $\overline{y} = y \cdot g^{\overline{x}}$ leads to the new most significant bit $\mathsf{msb}(x + \overline{x} \mod q)$. The latter is difficult to predict even if $x + \overline{x} < q$, since

$$\mathsf{msb}(x + \overline{x}) \begin{cases} x_k \oplus \overline{x}_k, & \text{if } x + \overline{x} < q \wedge x_{k-1} \dots x_0 + \overline{x}_{k-1} \dots \overline{x}_0 < 2^k \ , \\ 1 \oplus x_k \oplus \overline{x}_k, & \text{if } x + \overline{x} < q \wedge x_{k-1} \dots x_0 + \overline{x}_{k-1} \dots \overline{x}_0 \geq 2^k \end{cases}$$

and we have no information about the tail $x_{k-1} \dots x_0$.