Problem Set 4 Solutions

Problem 9. *In class we defined the multiquery PRG advantage for a PRG* $G\colon \{0,1\}^\ell \to \{0,1\}^L$ *by way of*

$$\mathbf{Adv}_G^{\mathrm{prg}*}(\mathcal{A}) = \Pr[\mathcal{A}^{\mathrm{G}} \Rightarrow 1] - \Pr[\mathcal{A}^{\$} \Rightarrow 1]$$

where the first oracle answers any query by $G(S)$*, for a freshly chosen* $S \twoheadleftarrow \{0,1\}^\ell$*, and the second oracle answers any query by returning a freshly chosen* $R \twoheadleftarrow \{0,1\}^L$*. Consider* $G = RC4$*, thought of as a map from 16 bytes to two (or more) bytes.*

Assume, as your experiments for Prob. 8 suggested, that the second byte of RC4 output is zero with probability $1/128$*. Design an adversary that breaks the security of RC4 with prg∗ advantage at least* 0.99*. For your analysis, you can use the following tool:*

Hoeffding's inequality. (See the Wikipedia entry with this name for more information.)
Let $X_1, \ldots, X_n$ *be independent and identically distributed random variables, each in* $\{0,1\}$ *and each taking on the value 1 with probability* p*. Let* $\overline{X} = \frac{1}{n}\sum X_i$ *be the "empirical mean" of the observations, which has the expected value of* $\mathrm{E}[\overline{X}] = p$*. Then for all real numbers* $t \geq 0$*,*

$$\Pr[|\overline{X} - p| \geq t] \leq 2e^{-2nt^2} .$$

Our adversary $\mathcal{A}$ will request n output samples of two bytes each, for a value n that we will determine from the analysis below. It will then compute the fraction of the time $\overline{X}$ that the second byte was 0. We are expecting this value either to be close to $1/128 = 4/512$ or close to $1/256 = 2/256$, so let's define $\mathcal{A}$ to output 1 if it observes $\overline{X} \geq 3/256$ and output 0 if it observes $\overline{X} < 3/256$.

Let $t = 1/513$. If $\overline{X}$ is in $[1/128 - t, 1/128 + t]$ then A will output 1. If $\overline{X}$ is in $[1/256 - t, 1/256 + t]$ then A will output 0. If $\overline{X}$ is in neither range, we don't care what it outputs.

Alternatively and more simply, we can have A answer 1 if $\overline{X} > 3/512$, and 0 otherwise, as this simplified algorithm complies with the mandated behavior above.

We now bound $\mathcal{A}$'s advantage as a function of n. Let X be the RV that is $\mathcal{A}$'s measurement when it speaks to the RC4 oracle, and let Y be the RV that is $\mathcal{A}$'s measurement when it speaks to the random-bits oracle. Then

$$\begin{aligned}
\mathbf{Adv}_{\mathrm{RC4}}^{\mathrm{prg}*}(\mathcal{A}) &= \Pr\left[\mathcal{A}^{\mathrm{RC4}(\cdot)} \Rightarrow 1\right] - \Pr\left[\mathcal{A}^{\$(\cdot)} \Rightarrow 1\right] \\
&= 1 - \Pr\left[\mathcal{A}^{\mathrm{RC4}(\cdot)} \Rightarrow 0\right] - \Pr\left[\mathcal{A}^{\$(\cdot)} \Rightarrow 1\right] \\
&\geq 1 - \Pr\left[\left|X - \frac{1}{128}\right| \geq \frac{1}{513}\right] - \Pr\left[\left|Y - \frac{1}{256}\right| \geq \frac{1}{513}\right] \\
&\geq 1 - 4e^{-2n(1/513)^2}
\end{aligned}$$

We seek adversarial advantage of at least $1 - 1/100$, so we should select n large enough that

$$4e^{-2n(1/513)^2} \quad \leq \quad \frac{1}{100}$$

or, solving for n, it suffices to have

$$n \geq \frac{513^2 \cdot \ln 400}{2} .$$

Google's calculator tells me that $n = 800,000$ suffices (rounding up to a nice round value). This is pretty striking: fewer than a million samples suffice for superb accuracy as to whether you're speaking to an RC4 generator or a generator of truly random bits.

The number n can be substantially lowered by switching to an appropriate (one-sided) Chernoff bound, which works better here. I did that in discussion section, ending up with $n \approx 21,000$.

Problem 10. *For this problem you will prove that PRG-security (the adversary is given one sample) is essentially equivalent to PRG*-security (where the adversary is given as many samples as it likes). More specifically:*

(a) *Let adversary $\mathcal{A}$ have advantage $\delta = \mathbf{Adv}_G^{\mathrm{prg}}(\mathcal{A})$ in attacking $G\colon\ \{0,1\}^\ell \to \{0,1\}^L$. Exhibit an adversary $\mathcal{B}$ of comparable efficiency that has "good" $\mathbf{Adv}_G^{\mathrm{prg*}}(\mathcal{B})$ advantage.*

This part is easy: $\mathcal{B}$ asks its oracle a single query, getting a response Y; then $\mathcal{B}$ runs $\mathcal{A}(Y)$, outputting what $\mathcal{A}$ does. Adversary $\mathcal{B}$'s behavior precisely emulates the defining behavior for $\mathcal{A}$'s, whence $\mathbf{Adv}_G^{\mathrm{prg*}}(\mathcal{B}) = \delta$. Of course $\mathcal{B}$ is efficient, asking a single query and running in approximately $\mathcal{A}$'s time

(b) *Let adversary $\mathcal{B}$ have advantage $\delta^* = \mathbf{Adv}_G^{\mathrm{prg*}}(\mathcal{B})$ in attacking $G\colon\ \{0,1\}^\ell \to \{0,1\}^L$. Exhibit an adversary $\mathcal{A}$ of comparable efficiency that has "good" $\mathbf{Adv}_G^{\mathrm{prg}}(\mathcal{A})$ advantage.*

The reduction is a hybrid argument. Let q be the maximum number of oracle queries asked by $\mathcal{B}$. Without loss of generality, assume that $\mathcal{B}$ always asks exactly q queries. (This entails no loss of generality insofar as $\mathcal{B}$ can always ask extra questions and ignore the answers.) We construct an adversary $\mathcal{A}$, approximately as efficient as $\mathcal{B}$, that, on input Y, gets advantage $\mathbf{Adv}_G^{\mathrm{prg}}(\mathcal{A}) = \delta^*/q$. Define:

> **algorithm** $\mathcal{A}(Y)$
> $j \twoheadleftarrow [1..q]$
> **for** $i \leftarrow 1$ to $j-1$ **do** $S_i \twoheadleftarrow \{0,1\}^\ell$, $Y_i \leftarrow G(S_i)$
> $S_j \twoheadleftarrow Y$
> **for** $i \leftarrow j+1$ to q **do** $Y_i \twoheadleftarrow \{0,1\}^L$
> Run $\mathcal{B}^{\mathcal{O}}$, answering $\mathcal{B}$'s ith query with Y_i and letting b be the $\mathcal{B}$'s final output
> **return** b

We observe that when $j = 1$ and $Y \twoheadleftarrow G(S)$ we are running $\mathcal{B}$ in an environment that corresponds to the experiment we denoted G (the first experiment in the definition of the adversary's advantage); and when $j = q$ and $Y \twoheadleftarrow \{0,1\}^L$ we are running $\mathcal{B}$ in an environment that corresponds to the experiment we denoted \$ (the second experiment in the definition of the adversary's advantage). By hybrid argument $\mathbf{Adv}_G^{\mathrm{prg}}(\mathcal{A}) = \delta/q$.

Problem 11. *On March 28 colleague Ross Anderson `https://www.cl.cam.ac.uk/~rja14/` died at his home in Cambridge, England. Read one or more papers by Anderson, and write a couple of pages in summary or analysis.*