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

Exercise (Difference between random self-reducibility reductions of CDH). *Let $\mathbb{G}$ be a finite group of prime order q such that all elements $y \in \mathbb{G}$ can be expressed as powers of $g \in \mathbb{G}$. Then the Computational Diffie-Hellman (CDH) problem is following. Given $x = g^a$ and $y = g^b$, find a group element $z = g^{ab}$. There are two well known ways to reduce an instance of CDH to a random instance of CDH*

```
$\mathcal{B}(x, y)$
    $r, s \leftarrow \mathbb{Z}_q$
    $w \leftarrow \mathcal{A}(x \cdot g^r, y \cdot g^s)$
    $z \leftarrow \dfrac{w}{x^s \cdot y^r \cdot g^{rs}}$
    return $z$
```

```
$\mathcal{C}(x, y)$
    $r, t \leftarrow \mathbb{Z}_q, s \leftarrow \mathbb{Z}_q^*$
    $w \leftarrow \mathcal{A}(x \cdot g^r, y^s \cdot g^t)$
    $z \leftarrow \left(\dfrac{w}{x^t \cdot y^{rs} \cdot g^{rt}}\right)^{-s}$
    return $z$
```

where $\mathcal{A}$ is the algorithm designed to solve a random instance of CDH. Both of these algorithms can be used to amplify the success probability by majority voting. Analyse the success of both strategies

```
$\mathcal{B}_n(x, y)$
    For $i \in \{1, \ldots, n\}$ do
        $r, s \leftarrow \mathbb{Z}_q$
        $w_i \leftarrow \mathcal{A}(x \cdot g^r, y \cdot g^s)$
        $z_i \leftarrow \dfrac{w_i}{x^s \cdot y^r \cdot g^{rs}}$
    return $\textsc{Majority}(z_1, \ldots, z_n)$
```

```
$\mathcal{C}_n(x, y)$
    For $i \in \{1, \ldots, n\}$ do
        $r, t \leftarrow \mathbb{Z}_q, s \leftarrow \mathbb{Z}_q^*$
        $w_i \leftarrow \mathcal{A}(x \cdot g^r, y^s \cdot g^t)$
        $z_i \leftarrow \left(\dfrac{w_i}{x^t \cdot y^{rs} \cdot g^{rt}}\right)^{-s}$
    return $\textsc{Majority}(z_1, \ldots, z_n)$
```

by defining a sharp enough lower bound on the success probability. Sketch how the bound behaves as a function of n. How big must $\mathsf{Adv}^{\mathsf{cdh}}_{\mathbb{G}}(\mathcal{A})$ be to get a noticeable success probability?

Solution. We divide the analysis into three distinct parts. First, we establish correctness. Second, we establish properties of the output distribution z. Third, we use established results together with standard probability bounds to lower and upper bound the success probability.

CORRECTNESS. Let α, β, γ be such that $x = g^\alpha$, $y = g^\beta$, $w = g^\gamma$ and $z = g^\delta$. Now if $\mathcal{A}$ returns a correct answer then $\gamma = (\alpha + r)(\beta + s)$ in the algorithm $\mathcal{B}$ and $\gamma = (\alpha + r)(\beta s + t)$ in the algorithm $\mathcal{C}$. Consequently,

$$\delta = (\alpha + r)(\beta + s) - \alpha s - \beta r - rs = \alpha\beta$$

in the algorithm $\mathcal{B}$ and

$$\delta s = (\alpha + r)(\beta s + t) - \alpha t - \beta rs - rt = \alpha\beta s$$

in the algorithm $\mathcal{C}$ and the reconstruction of z is indeed successful.

ANALYSIS OF OUTPUT DISTRIBUTION. First, note that additive noise component in both algorithms completely blinds inputs and thus $\mathcal{A}$ receives a CDH random instance. The difference in the algorithm emerges if $\mathcal{B}$ gives a wrong answer. Let us assume that the answer of $\mathcal{A}$ is offset by term g^ϵ. Then the same offset propagates to the final answer in the algorithm $\mathcal{B}$ as

$$\delta = (\alpha + r)(\beta + s) + \epsilon - \alpha s - \beta r - rs = \alpha\beta + \epsilon \ .$$

In algorithm $\mathcal{C}$, the offset is randomised as

$$\delta s = (\alpha + r)(\beta s + t) + \epsilon - \alpha t - \beta rs - rt = \alpha\beta s + \epsilon \ .$$

Since $\delta = \alpha\beta + \epsilon s^{-1}$ and s is chosen uniformly from $\mathbb{Z}_q^*$, we get that δ is also uniformly distributed over $\mathbb{Z}_q \setminus \{\alpha\beta\}$ as the group size q is a prime number by our assumptions. When the group size is not prime, the term ϵs^{-1} may cycle through a proper subgroup of $\mathbb{Z}_q$ and thus the wrong answer is not converted into $q-1$ equiprobable values of z rather there are fewer but still remarkable number of equiprobable values of z. The following allows to generalise the following results for groups where the order of group is a composite number but amplification is less powerful but such drop in efficiency is unavoidable.

Success bounds for simple amplification strategy. Let us now analyse the algorithm $\mathcal{B}_n$. The worst case for majority voting is where queries to $\mathcal{A}$ gives either the correct answer or a wrong answer with constant offset ϵ. Then the correct answer competes against single wrong answer and amplification is possible only if the probability of a correct answer $\mathsf{Adv}^{\mathsf{cdh}}_{\mathbb{G}}(\mathcal{A}) > \frac{1}{2}$ Then we can encode the correct answer as 1 and the wrong answer as 0 and view the majority voting process as a binomial distribution which yields 1 with probability $\mathsf{Adv}^{\mathsf{cdh}}_{\mathbb{G}}(\mathcal{A}) = \varepsilon$ and 0 with probability $1-\varepsilon$. The majority voting then yields the right answer if the sum of the trials is greater than $\frac{n}{2}$. Let us recall that Hoeffding's inequality implies

$$\Pr\left[\sum_{i=1}^{n} X_i \le (\varepsilon - \rho)n\right] \le e^{-2\rho^2 n}$$

for independently drawn indicator variables $X_i \in \{0,1\}$ such that $\Pr\left[X_i = 1\right] = \varepsilon$. Consequently, if we encode success of an i-th invocation of $\mathcal{A}$ as the indicator X_i, we obtain an upper bound on the failure probability

$$\Pr\left[\mathcal{B}_n(g^\alpha, g^\beta) \neq g^{\alpha\beta}\right] = \Pr\left[\sum_{i=1}^{n} X_i \le \frac{n}{2}\right] \le e^{-\frac{n(2\varepsilon-1)^2}{2}}$$

that decreases exponentially wrt n. Note that the bound becomes effective only if $\varepsilon > \frac{1}{2}$. The same argument shows that amplification quickly converges to a wrong answer in the worst case scenario where $\mathcal{A}$ always has a fixed offset δ for wrong answers and thus $\mathcal{B}_n$ is bound to fail when $\varepsilon < \frac{1}{2}$.

Success bounds for complex amplification strategy. Recall that by construction z_i is a uniformly chosen among all possible answers that are wrong when $\mathcal{A}$ fails. As a result, we can express

$$\Pr\left[z_i = g^{\alpha\beta+\epsilon}\right] = \begin{cases} \varepsilon & \text{if } \epsilon = 0\ , \\ \frac{1-\varepsilon}{q-1} & \text{if } \epsilon \neq 0\ . \end{cases}$$

As the majority voting amplifies the probability of the most probable answer, the algorithm is applicable only if the probability of z_i being correct is larger than a probability of any wrong value. From which we can conclude $\varepsilon > \frac{1}{q}$, i.e., when $\mathcal{A}$ must work better than random guessing.

Now note that correct answer can win only if the maximal number of wrong answers is at least one smaller than the number of correct answers. Let $F_{q-1,k}(\ell)$ denote the probability that when we sample k values uniformly form $\mathbb{Z}_q^*$ maximal count of identical values is strictly smaller than ℓ. Then we can decompose the probability of a correct answer as follows

$$\Pr\left[\mathcal{C}_n(g^\alpha, g^\beta) = g^{\alpha\beta}\right] = \sum_{\ell=2}^{n} \binom{n}{\ell} \varepsilon^\ell (1-\varepsilon)^{n-\ell} \cdot F_{q-1,n-\ell}(\ell)$$

as the first factor in the term corresponds to the probability that we obtain exactly ℓ correct answers and the second term corresponds to the probability that the remaining $n-\ell$ wrong answers do not overthrow the right verdict. For reasonable parameter values, the first term in the sum is the largest and thus we get a quite effective lower bound on the success:

$$\Pr\left[\mathcal{C}_n(g^\alpha, g^\beta) = g^{\alpha\beta}\right] \ge \frac{n(n-1)\varepsilon^2(1-\varepsilon)^{n-2}}{2} \cdot F_{q-1,n-2}(2)$$

where the second probability $F_{q-1,n-2}(2)$ corresponds to the generalised birthday problem.

[Off by 1-F]

When the number of wrong answers $n-2 \le \sqrt{q-1}$ then the corresponding lower and upper bounds are quite tight:

$$0.316 \cdot \frac{(n-2)(n-3)}{q-1} \le F_{q-1,k}(2) \le 0.5 \cdot \frac{(n-2)(n-3)}{q-1} \quad .$$

As there are generic algorithms for finding discrete logarithm in time $\Theta(\sqrt{q})$, the behaviour of the amplification algorithm is uninteresting for $n \ge \sqrt{q}$ and we do not have to consider bounds for the other ranges. Consequently, we can conclude

$$\Pr\left[\mathcal{C}_n(g^{\alpha}, g^{\beta}) = g^{\alpha\beta}\right] \ge 0.158 \cdot \frac{n(n-1)(n-2)(n-3)\varepsilon^2(1-\varepsilon)^{n-2}}{q-1}.$$

The probability that all $n-2$ draws are distinct elements can

To find the probability that at least two numbers are the same, we can use the findings of the generalized birthday problem.[1] It has been shown that when drawing n' random integers from a discrete uniform distribution with range $[1,d]$, the probability $P(n',d)$ that at least two numbers are the same is

$$P(n',d) \approx 1 - e^{\frac{-n'(n'-1)}{2d}}$$

which in our case would be

$$P(n-2, |\mathbb{G}|-1) \approx 1 - e^{\frac{-(n-2)(n-3)}{2(|\mathbb{G}|-1)}}$$

Thus $P(A \cap B) \approx (\frac{n(n-1)}{2}\varepsilon^2(1-\varepsilon)^{n-2})(1-e^{\frac{-(n-2)(n-3)}{2(|\mathbb{G}|-1)}})$ And the probability of $\mathcal{C}_n$ succeeding in this game is

$$P(A \setminus B) = P(A) - P(A \cap B) \approx \tfrac{n(n-1)}{2}\varepsilon^2(1-\varepsilon)^{n-2} - (\tfrac{n(n-1)}{2}\varepsilon^2(1-\varepsilon)^{n-2})(1-e^{\frac{-(n-2)(n-3)}{2(|\mathbb{G}|-1)}}) = \tfrac{n(n-1)}{2}\varepsilon^2(1-\varepsilon)^{n-2}e^{\frac{-(n-2)(n-3)}{2(|\mathbb{G}|-1)}}$$

Which gives us an approximate lower bound on the success of majority voting for $\mathcal{C}_n$.

??

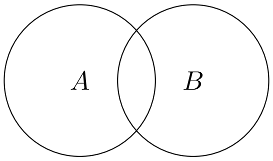
have less than ℓ coinciding values.

us denote the probability

$$F_n(\ell) = \Pr\left[x??\right]$$

??

Let us first view a different game where instead of doing majority voting, $\mathcal{C}_n$ succeeds if he gets two right answers and there are no coinciding wrong answers. Note that the probability of success in this game is a lower bound on the probability of success in majority voting, as in majority voting $\mathcal{C}_n$ has more win conditions (for example 3 correct answers and 2 coinciding wrong answers). The probability that $\mathcal{C}_n$ outputs exactly two right answers is $\binom{n}{2}\varepsilon^2(1-\varepsilon)^{n-2} = \frac{n(n-1)}{2}\varepsilon^2(1-\varepsilon)^{n-2}$. In the Venn diagram below, this is our probability of getting A, where we draw two right answers. B stands for drawing two coinciding wrong answers. Our new game wins if we get $A \setminus B$. Thus we have to subtract the probability of getting $A \cap B$.



To find $P(A \cap B)$ we can use the conditional probability formula $P(A \cap B) = P(A)P(B|A)$. $P(A) = \frac{n(n-1)}{2}\varepsilon^2(1-\varepsilon)^{n-2}$ as above. $P(B|A)$ however stands for drawing at least two coinciding wrong answers, conditioned that we have drawn exactly two right answers. Since in this condition we have drawn two right answers, we have sampled the wrong answer distribution $n-2$ times.

[1] https://en.wikipedia.org/wiki/Birthday_problem#Cast_as_a_collision_problem

??

Example: Let's view the probability of winning this game in a setting where $\mathbb{G}$ contains all 100 numbers and $\mathsf{Adv}^{\mathsf{cdh}}_{\mathbb{G}}(\mathcal{A}) = \varepsilon = 0.2$. The probability of a fixed incorrect answer is then $\frac{0.8}{100-1}$. Figure 1 shows the approximate success probability of winning this game for a small range of n's. We can see that at first we do get some amplification, for $n = 8$, the probability of success is around 0.25 which is an increase compared to $\varepsilon = 0.2$. However after that the success probability drops off. The reason for that is that the probability of two wrong answers coinciding is increasing significantly in that region, which causes loss in the game. However, in the majority voting case there will still be amplification going on, but to get a more accurate estimate of the amplification, we would have to view also the cases of reaching 3 or 4 correct answers with less coinciding wrong answers.

Figure 1: Approximate probability of $\mathcal{C}_n$ succeeding in the game, $\varepsilon = 0.2$, $|\mathbb{G}| = 100$

??

For example, if $p = \mathsf{Adv}^{\mathsf{cdh}}_{\mathbb{G}}(\mathcal{A}) = 0.55$, then $\varepsilon = 0.05$. Figure 2 illustrates lower bound on the probability that majority voting succeeds on a small range of n's. Note that by using this bound we can compute the n needed to find as high of a success probability as we want.

Figure 2: Lower bound on success ratio in given example $p = 0.55$, $\varepsilon = 0.05$