Karp Reduction Of Private Coin HVSZK Protocol To Entropy Difference

Personal Notes

Ari Biswas
University Of Warwick
aribiswas3@gmail.com

1 Prelims

Definition 1 (Entropy) If X is a discrete probability distribution, then the entropy of X, denoted $\mathsf{H}(X)$, is defined as

$$\mathsf{H}(X) = \sum_{x} \Pr[X = x] \cdot \log \frac{1}{\Pr[X = x]} = \mathbb{E}_{x \xleftarrow{\$} X} \left[\log \frac{1}{\Pr[X = x]} \right]$$

Definition 2 (KL Divergence) Let X and Y be two discrete distributions. The relative entropy (or KL Divergence) between X and Y is given by

$$\mathsf{KL}\left(X, Y\right) = \mathbb{E}_{\alpha \xleftarrow{\$} X} \left[\log \frac{\Pr[X = \alpha]}{\Pr[Y = \alpha]} \right]$$

2 Goal

We have some promise problem $\Pi = (\Pi^{\mathsf{Yes}}, \Pi^{\mathsf{No}})$ and a private coin HVSZK proof system (P, V) for Π with completeness and soundness error of 2^{-40}. Let Sim denote the simulator for the proof system described above. Given $x \in \Pi^{\mathsf{Yes}} \cup \Pi^{\mathsf{No}}$, define the security parameter as $\kappa = |x|$. We want to use the output of the simulator $\mathsf{Sim}(1^{\kappa}, x)$[1] to come up with two distributions X and Y such that

$$x \in \Pi^{\mathsf{Yes}} \implies \mathsf{H}(X) \geq \mathsf{H}(Y) + 1$$

$$x \in \Pi^{\mathsf{No}} \implies \mathsf{H}(Y) \geq \mathsf{H}(X) + 1$$

where $\mathsf{H}(X)$ and $\mathsf{H}(Y)$ refers to the entropy of distributions X and Y as defined in Definition 1.

3 Notation

Let (P, V) be an interactive HVSZK proof system for a promise problem Π. Let Sim be the simulator. Let $v(\kappa)$ denote the poly bound on the number of messages V sends to the prover. Let $t(\kappa)$ be the poly bound on the total communication over the proof in bits, and $r(\kappa)$ denote the number of random bits accessed during the proof by the verifier. Without loss of generality, assume that in the last round of the protocol, when the

[1] Henceforth, I'll often drop the security parameter from the notation to make it cleaner and write $\mathsf{Sim}(1^{\kappa}, x)$ as just $\mathsf{Sim}(x)$.

verifier outputs accept or reject, it also sends over all $r(\kappa)$ bits of randomness to the prover[2]. We will often abbreviations $r = r(\kappa), t = t(\kappa)$ and $v = v(\kappa)$.

Just like in the public coin HVSZK to statistical closeness reduction, we want to use the output distribution of the simulator to define distributions that allow us to differentiate between Yes and No instances. Part of this was quantifying the difference in behaviour between a virtual and a real verifier. In the public coin setting, we quantified the difference by the statistical difference between messages sent by the virtual verifier and an algorithm that picked messages uniformly at random, independent of message history (like the prescribed verifier). We cannot use such a prescribed verifier this time as the verifier's coins are not known in advance. Thus, we must quantify the difference between the virtual and real verifiers differently. The following technical lemmas help us do precisely this.

4 Technical Lemmas

Note that $\mathsf{H}(\mathsf{Sim}(x)_{2_i}|\mathsf{Sim}(x)_{2_i-1}) = \mathsf{H}(\mathsf{Sim}(x)_{2_i}) - \mathsf{H}(\mathsf{Sim}(x)_{2_i-1})$. It measures how many bits of randomness the i'th message of the virtual verifier contributes to the output distribution of the simulator. Since the real verifier uses $r(\kappa)$ bits of randomness over the whole proof, the sum of these terms should be close to $r(\kappa)$ if the simulation is good.

Lemma 1

$$\begin{aligned} \mathsf{KL}\left(\mathsf{Sim}(x), (\mathsf{P}_{\mathsf{Sim}}, \mathsf{V})(x)]\right) &= r(\kappa) - \sum_{i=1}^{v(\kappa)+1} \mathsf{H}(\mathsf{Sim}(x)_{2_i}|\mathsf{Sim}(x)_{2_i-1}) && (1) \\ &= r(\kappa) - \sum_{i=1}^{v(\kappa)+1} \mathsf{H}(\mathsf{Sim}(x)_{2_i}) - \mathsf{H}(\mathsf{Sim}(x)_{2_i-1}) && (2) \end{aligned}$$

Proof. For γ let γ_i denote the first i messages.

$$\begin{aligned} \mathsf{KL}\left(\mathsf{Sim}, (\mathsf{P}_{\mathsf{Sim}}, \mathsf{V})\right) &= \sum_{\gamma} \Pr[\mathsf{Sim} = \gamma] \cdot \log \frac{\Pr[\mathsf{Sim} = \gamma]}{\Pr[(\mathsf{P}_{\mathsf{Sim}}, \mathsf{V}) = \gamma]} && (3) \\ &= \sum_{\gamma} \Pr[\mathsf{Sim} = \gamma] \cdot \log \frac{\prod_{i=1}^{2v} \Pr[\mathsf{Sim}_i = \gamma_i | \mathsf{Sim}_{i-1} = \gamma_{i-1}]}{\prod_{i=1}^{2v} \Pr[(\mathsf{P}_{\mathsf{Sim}}, \mathsf{V})_i = \gamma_i | (\mathsf{P}_{\mathsf{Sim}}, \mathsf{V})_{i-1} = \gamma_{i-1}]} && (4) \\ &= \sum_{\gamma} \Pr[\mathsf{Sim} = \gamma] \cdot \log \frac{\prod_{i=1}^{v} \Pr[\mathsf{Sim}_{2i} = \gamma_{2i} | \mathsf{Sim}_{2i-1} = \gamma_{2i-1}]}{\prod_{i=1}^{v} \Pr[(\mathsf{P}_{\mathsf{Sim}}, \mathsf{V})_{2i} = \gamma_{2i} | (\mathsf{P}_{\mathsf{Sim}}, \mathsf{V})_{2i-1} = \gamma_{2i-1}]} && (5) \end{aligned}$$

Where the last inequality is by the definition of the virtual prover. $\mathsf{P}_{\mathsf{Sim}}$ is just the simulator pretending to be the real prover.

Note that for any γ, the denominator of the above fraction equals the reciprocal of the number of possible outcomes of the verifier coins 2^{-r}, since the even indexed messages of $(\mathsf{P}_{\mathsf{Sim}}, \mathsf{V})$ are generated exactly as the real verifier would. And we know the real verifier uses r bits of randomness.

[2] This does not affect soundness as the verifier only does this in the round in which it makes its decision. A cheating prover cannot use this randomness to its advantage.

$$\begin{align}
\mathsf{KL}\left(\mathsf{Sim},(\mathsf{P}_{\mathsf{Sim}},\mathsf{V})\right) &= \sum_{\gamma} \Pr[\mathsf{Sim}=\gamma]\cdot\log\frac{\prod_{i=1}^{v}\Pr[\mathsf{Sim}_{2i}=\gamma_{2i}|\mathsf{Sim}_{2i-1}=\gamma_{2i-1}]}{\prod_{i=1}^{v}\Pr[(\mathsf{P}_{\mathsf{Sim}},\mathsf{V})_{2i}=\gamma_{2i}|(\mathsf{P}_{\mathsf{Sim}},\mathsf{V})_{2i-1}=\gamma_{2i-1}]} \tag{6}\\
&= \sum_{\gamma} \Pr[\mathsf{Sim}=\gamma]\cdot\log\frac{\prod_{i=1}^{v}\Pr[\mathsf{Sim}_{2i}=\gamma_{2i}|\mathsf{Sim}_{2i-1}=\gamma_{2i-1}]}{2^{-r}} \tag{7}\\
&= \sum_{\gamma} \Pr[\mathsf{Sim}=\gamma]\cdot\log\frac{\prod_{i=1}^{v}\Pr[\mathsf{Sim}_{2i}=\gamma_{2i}|\mathsf{Sim}_{2i-1}=\gamma_{2i-1}]\cdot\prod_{i=1}^{v}\Pr[\mathsf{Sim}_{2i-1}=\gamma_{2i-1}]}{2^{-r}\cdot\prod_{i=1}^{v}\Pr[\mathsf{Sim}_{2i-1}=\gamma_{2i-1}]} \tag{8}\\
&= \sum_{\gamma} \Pr[\mathsf{Sim}=\gamma]\cdot\log\frac{\prod_{i=1}^{v}\Pr[\mathsf{Sim}_{2i}=\gamma_{2i}]}{2^{-r}\cdot\prod_{i=1}^{v}\Pr[\mathsf{Sim}_{2i-1}=\gamma_{2i-1}]} \tag{9}\\
&= \sum_{i=1}^{v}\sum_{\gamma}\Pr[\mathsf{Sim}=\gamma]\cdot\Pr[\mathsf{Sim}_{2i}=\gamma_{2i}]+r+\sum_{i=1}^{v}\sum_{\gamma}\Pr[\mathsf{Sim}=\gamma]\cdot\Pr[\mathsf{Sim}_{2i-1}=\gamma_{2i-1}] \tag{10}\\
&= \sum_{i=1}^{v}\mathsf{H}(\mathsf{Sim}_{2_j})+r+\sum_{i=1}^{v}\mathsf{H}(\mathsf{Sim}_{2_j-1}) \tag{11}
\end{align}$$

□

Fact 1 For any two random variables X and Y, ranging over a common universe Ω and $\delta = d_{\mathtt{TV}}(X,Y)$ we have

$$|\mathsf{H}(X)-\mathsf{H}(Y)| \le \log(|\Omega|-1)\cdot\delta+\mathsf{H}_2(\delta) \tag{12}$$

See any information theory textbook for proof using Fano's inequality. Vadhan does provide a more direct, elegant proof.

Lemma 2 Let $\delta(x) = d_{\mathtt{TV}}(\mathsf{Sim}(x), \mathrm{view}[(\mathsf{P},\mathsf{V})(x)])$. Then,

$$r(\kappa) - \sum_{i=1}^{v(\kappa)+1}\mathsf{H}(\mathsf{Sim}(x)_{2_i}|\mathsf{Sim}(x)_{2_i-1}) \le 2(v(\kappa)+1)\left[t'(x)\delta(x)+\mathsf{H}_2(\delta(x))\right] \tag{13}$$

Proof. Consider a perfect simulator $\overline{\mathsf{Sim}}$ with 0 deviation for (P,V). The simulation based prover with respect to $\overline{\mathsf{Sim}}$ is simply P. By Lemma 1 we have

$$\begin{align}
r+\sum_{i=1}^{2(v+1)}(-1)^{i+1}\cdot\mathsf{H}((\mathsf{P},\mathsf{V})_i) &= r+\sum_{i=1}^{2(v+1)}(-1)^{i+1}\cdot\mathsf{H}(\overline{\mathsf{Sim}_i}) \tag{14}\\
&= \mathsf{KL}\left(\overline{\mathsf{Sim}},(\mathsf{P},\mathsf{V})\right) \tag{15}\\
&= 0 \tag{16}
\end{align}$$

Equation (14) is just a compact way to represent the difference terms in Lemma 1 and the next line follows from the lemma. We get that it's 0 by assumption of perfect simulation. Now we have

$$\begin{aligned} r+\sum_{i=1}^{2(v+1)}(-1)^{i+1}\cdot\mathsf{H}(\mathsf{Sim}_i) &\le r+\sum_{i=1}^{2(v+1)}(-1)^{i+1}\cdot\mathsf{H}((\mathsf{P},\mathsf{V})_i)+\sum_{i=1}^{2(v+1)}|\mathsf{H}(\mathsf{Sim}_i)-\mathsf{H}((\mathsf{P},\mathsf{V}))| && (17)\\ &= 0+\sum_{i=1}^{2(v+1)}|\mathsf{H}(\mathsf{Sim}_i)-\mathsf{H}((\mathsf{P},\mathsf{V}))| && (18)\\ &\le 2(v+1)\cdot[\delta(t+1)+\mathsf{KL}_2\left(\delta\right)] && (19)\end{aligned}$$

The last inequality comes from Fact 1.

□

Fact 2 For any two distributions X and Y, we have

1. $\mathsf{KL}\left(X,Y\right)\ge 0$, and $\mathsf{KL}\left(X,Y\right)=0 \iff X=Y$
2. For any function f, $\mathsf{KL}\left(f(X),f(Y)\right)\le\mathsf{KL}\left(X,Y\right)$.
3. For $0\le q'\le q\le p\le p'\le 1$, $\mathsf{KL}_2\left(p',q'\right)\ge\mathsf{KL}_2\left(p,q\right)$, where $\mathsf{KL}_2\left(\cdot,\cdot\right)$ is the KL distribution between Bernoulli distributions.

Fact 3 For every joint distribution (X,Y),

1. $\mathsf{H}(X|Y)\le\mathsf{H}(X)$
2. $\mathsf{H}(X,Y)=\mathsf{H}(Y)+\mathsf{H}(X|Y)$

where $\mathsf{H}(X|Y)=\mathbb{E}_{y\overset{\$}{\leftarrow}Y}[\mathsf{H}(X|y)]$

Lemma 3 Let p denote the simulators acceptance probability, and q be the maximum taken over all provers $\widetilde{\mathsf{P}}$, such that V accepts. Assuming $p\ge q$, then

$$\mathsf{KL}_2\left(p,q\right)\le\mathsf{KL}\left(\mathsf{Sim}(x),\mathrm{view}[(\mathsf{P}_{\mathsf{Sim}},\mathsf{V})(x)]\right) \tag{20}$$

Proof. Let f be some boolean function such that $f(\gamma)=1$ if γ is accepting (by the real verifier) and 0, otherwise. By Fact 2,

$$\mathsf{KL}\left(\mathsf{Sim},(\mathsf{P}_{\mathsf{Sim}},\mathsf{V})\right)\ge\mathsf{KL}\left(f(\mathsf{Sim}),f((\mathsf{P}_{\mathsf{Sim}},\mathsf{V}))\right)=\mathsf{KL}_2\left(p,q'\right)=\mathsf{KL}_2\left(p,q\right) \tag{21}$$

where $q'\le q$ equals the probability of the real verifier accepting when interacting with the $(\mathsf{P}_{\mathsf{Sim}},\mathsf{V})$.

□

5 The Reduction

Combining all the lemmas above, we get

$$\mathsf{KL}_2\left(p,q\right)\le 2(v(\kappa)+1)\left[t'(x)\delta(x)+\mathsf{H}_2(\delta(x))\right]$$

Theorem 1 Every promise problem with a weak public coin honest verifier statistical zero knowledge proof reduces to ED.

Proof. Assume Soundness and Completeness Errors $\delta_c = \delta_s = 2^{-40}$ and simulator deviation $\mu \leq \min\{\frac{1}{v't'}, \varepsilon\}$[3], where ε is to be defined. Given $x \in \Pi$ and the Simulator Sim we want to construct X and Y as follows:

Define $X = \mathsf{Sim}(x)_2 \otimes \mathsf{Sim}(x)_4 \ldots, \otimes \mathsf{Sim}(x)_{2v'}$

Define $Y_1 = \mathsf{Sim}(x)_1 \otimes \mathsf{Sim}(x)_3 \ldots, \otimes \mathsf{Sim}(x)_{2v'-1}$
Define $Y_2 \overset{\$}{\leftarrow} \{0,1\}^{r-7}$
Define Y_3: Run the simulator $8\log(t'v'+2)$ times independently. If the virtual verifier rejects in the majority of the transcripts, set $Y_3 \overset{\$}{\leftarrow} \{0,1\}^{t'v'+2}$. Otherwise, set $Y_3 = \perp$.
Define $Y = Y_1 \otimes Y_2 \otimes Y_3$.

Yes Case Now if $x \in \Pi^{\mathsf{Yes}}$, then by Lemma 2, we have

$$r - \mathsf{H}(X) + \mathsf{H}(Y_1) \leq 2v'[t'\mu + \mathsf{H}(\mu)] \tag{22}$$

For very small values of δ such as $\delta < 0.01$, we have $\mathsf{H}_2(\delta) \leq \sqrt{\delta}$, therefore setting $\varepsilon = 0.01$, we get $\mathsf{H}(\mu) \leq \sqrt{\mu}$ and as by the simulation assumption we have $\mu \leq \frac{1}{v't'}$ we have

$$\begin{aligned} r - \mathsf{H}(X) + \mathsf{H}(Y_1) &\leq 2v'[t'\mu + \mathsf{H}(\mu)] & (23)\\ &\leq 2 + 2\sqrt{\frac{1}{v't'}} & (24)\\ &\leq 4 & (25)\\ \mathsf{H}(Y_1) &\leq \mathsf{H}(X) + 4 - r & (26)\end{aligned}$$

By definition of Y_2, $\mathsf{H}(Y_2) = r - 7$, so we have

$$\mathsf{H}(Y_1) + \mathsf{H}(Y_2) \leq \mathsf{H}(X) + 4 - r + r - 7 \tag{27}$$

If we show that $\mathsf{H}(Y_3) \leq 2$, then we get

$$\mathsf{H}(Y) = \mathsf{H}(Y_1) + \mathsf{H}(Y_2) + \mathsf{H}(Y_3) \leq \mathsf{H}(X) - 1$$

and we are done. Let $A_1, \ldots, A_w$ be the different runs of the simulator where $w = 8\log(t'v'+2)$. As $x \in \Pi^{\mathsf{Yes}}$, we have the probability p of the simulator rejecting to be upper bounded as $p - \delta_c \leq \mu$ or $p \leq \mu + 2^{-40}$. Given that $\varepsilon \leq 0.01$ and $\mu \leq \varepsilon$ by definition, then it's safe to say that $p \leq \frac{1}{4}$. Now the probability of Y_3 outputting $t'v'+2$ random bits is the same as the probability $\Pr[\frac{1}{w}\sum_{i=1}^{w} A_i \leq 1/2] = \Pr[\frac{1}{w}\sum_{i=1}^{w} A_i - 1/4 \leq 1/4]$, which the Chernoff bound is at most $\frac{1}{t'v'+2}$. Let A be a Bernoulli random variable such that if $A = 1$, then Y_3 does not output $\perp$ and $\Pr[A = 1] = p' \leq \frac{1}{t'v'+2}$.

[3] Typically we assume deviation to be negligible so this a weaker assumption that ZK

$$\begin{aligned} \mathsf{H}(Y_3) &= \mathsf{H}(A, p') && (28)\\ &= \mathsf{H}(p') + \mathsf{H}(A|p') && (29)\\ &\leq 1 + p'(t'v' + 2) + (1 - p') \cdot 0 && (30)\\ &\leq 1 + 1 = 2 && (31) \end{aligned}$$

This handles the Yes case.

No Case But we still have to handle the No cases. We want to show that

$$x \in \Pi^{\mathsf{No}} \implies \mathsf{H}(Y) \geq \mathsf{H}(X) + 1$$

It suffices to show that either $x \in \Pi^{\mathsf{No}} \implies \mathsf{H}(X) \leq \mathsf{H}(Y_3) + 1$ or $x \in \Pi^{\mathsf{No}} \implies \mathsf{H}(X) \leq \mathsf{H}(Y_1) + \mathsf{H}(Y_2) + 1$. Assume that the simulator outputs accepting transcripts (virtual verifier accepts) with low probability at most $\frac{1}{4}$ (Low Acceptance For The Virtual Verifier). Once again by the Chernoff Bound, we will get the probability p with which the Simulator outputs $\perp$ is $p \leq \frac{1}{t'v'+2}$. Using the same analysis for $\mathsf{H}(Y_3)$ as above, we get

$$\begin{aligned} \mathsf{H}(Y_3) &\geq (1 - p')(t'v' + 2) && (32)\\ &\geq (t'v' + 1) && (33)\\ &\leq \mathsf{H}(X) + 1 && (34) \end{aligned}$$

(33) Comes from the fact that $p \leq \frac{1}{t'v'+2}$ so setting $p = \frac{1}{t'v'+2}$ gives us what we want and (34) comes from the fact that X includes at most $t'v'$ random bits.

Now suppose the simulator outputs accept with at least $\frac{1}{4}$ (High Acceptance For The Virtual Verifier), then by lemma 3 we have

$$\mathsf{KL}\left(\mathsf{Sim}(x), \mathrm{view}[(\mathsf{P}_{\mathsf{Sim}}, \mathsf{V})(x)]\right) \geq \mathsf{KL}_2\left(1/4, 2^{-40}\right) > 8$$

Now by Lemma 1 we have

$$\begin{aligned} r - \mathsf{H}(X) + \mathsf{H}(Y_1) &\geq 8 && (35)\\ \mathsf{H}(X) - \mathsf{H}(Y_1) &\leq r - 8 && (36)\\ \mathsf{H}(X) - \mathsf{H}(Y_1) - \mathsf{H}(Y_2) &\leq r - 8 + \mathsf{H}(Y_1) - \mathsf{H}(Y_2) && (37)\\ &\leq r - 8 + (r - 7) && (38)\\ \mathsf{H}(X) &\leq \mathsf{H}(Y_1) + \mathsf{H}(Y_2) + 1 && (39) \end{aligned}$$

□

References