INF581: Du calcul probabiliste au calcul quantique **Hiver 2011**

Cours 5 — 7 février

Enseignant : David Xiao *Rédacteur : Alba Marina MÁLAGA SABOGAL*

5.1 Motivation

La dernière fois, on a vu la définition d'une preuve interactive sous forme de dialogue. On peut légitimement se demander : Qu'est-ce que divulguent les preuves ? C'est à dire, une fois le dialogue entre le prouveur et le vérificateur fini, est-il possible que l'un ou l'autre ait obtenu de nouvelles informations ?

Par exemple, une preuve pour un langage dans la classe NP est une preuve fixe, donc après un premier dialogue, le verificateur peut divulguer la preuve, ce qui n'est à priori pas désirable, en tout cas ce n'est pas toujours acceptable

Figure 5.1. Divulgation de l'information dans une preuve NP

$P \overset{\pi}{\rightarrow} V \quad \Rightarrow$ Un vérificateur qui voulait juste être convaincu de l'appartenance de x au langage a réçu en plus le certificat qui confirme cette appartenance.

EXEMPLE : Le prouveur **(P)** est une entreprise avec une solution à un problème d'optimisation Le vérificateur **(V)** est un client. **(P)** veut convaincre **(V)** qu'il a la solution sans la reveler.

Ce qu'on va voir aujourd'hui est un peu plus ambitieux que ça. Le but de la leçon d'aujourd'hui est de comprendre comment dans une preuve interactive **(P)** convainque **(V)** sans rien reveler sauf la veracite de l'enonce.

5.2 Divulgation nulle à vérificateur honnête

Mais comment exprimer que **(V)** n'apprend rien ? En quelque sorte on voudrait que **(P)** dise seulement des choses que **(V)** connait déjà... donc informellement une preuve interactive serait ZK (zéro knowledge) si pour tout x dans le langage, **(V)** peut engendrer une preuve lui-même.

Si un langage $\mathcal{L}$ a à la fois une preuve non-intéractive et "ZK" alors $\mathcal{L} \in P$. Donc cette definition est seulement interessante si on regarde les preuves interactives avec de l'alea.

On va donner la definition en deux temps, d'abord la definition un peu faible et puis un peu forte.

Définition 5.1 (Point de vue). *Le point de view du verificateur dans une preuve interactive*

$$\begin{array}{cc} P & V \quad \omega \\ & \overset{q_1}{\longleftarrow} \\ & \overset{r_1}{\longrightarrow} \\ & \overset{q_2}{\longleftarrow} \\ & \overset{r_2}{\longrightarrow} \\ & \vdots \\ & \overset{q_m}{\longleftarrow} \\ & \overset{r_m}{\longrightarrow} \end{array}$$

est

$$VIEW_{P,V}(x) = (x, \omega, r_1, \ldots, r_m) \tag{5.1}$$

Définition 5.2 (Distance statistique). *La distance statistique entre X,Y deux variables aleatoires sur U*

$$\Delta(X, Y) = \max_{T \subseteq U} |Pr[X \in T] - Pr[Y \in T]| \tag{5.2}$$

On peut montrer que

$$\Delta(X, Y) = \frac{1}{2} \sum_{u \in U} |Pr[X = u] - Pr[Y = u]| \tag{5.3}$$

Définition 5.3 (Indistinguibilité statistique). *Soient $\{X_z\}_{z \in \{0,1\}^*}$, $\{Y_z\}_{z \in \{0,1\}^*}$ deux familles de variables aléatoires sur $\{U_z\}_{z \in \{0,1\}^*}$. $\{X_z\}$ et $\{Y_z\}$ sont statistiquement indistinguibles si pour tout $c > 0$, il existe un entier positif n_c tel que*

$$\forall |z| \geq n_c, \Delta(X_z, Y_z) \leq |z|^{-c}$$

Souvent, on demontre l'indistinguibilité statistique avec la borne exponentiellement petite $2^{-|z|}$

Définition 5.4 (HVZK : Zero knowledge(divulgation nulle) avec (V) honnête). *Un protocole **P,V** pour un langage $\mathcal{L}$ est HVZK si il existe **S** un "simulateur" efficace probabiliste tel que*

$$\forall x \in \mathcal{L}, S(x) \text{ et } VIEW(x) \text{ sont statistiquement indistinguables}$$

(Un "simulateur" est en fait une fonction qui à chaque x associe une variable aléatoire)

Cela formalise l'idée qu'on avait avant, c'est que le verificateur peut donner une preuve en utilisant le simulateur.

On revient à l'exemple du langage des couples de graphes non isomorphes vu la dernière fois.

$$GNI = \{(G_0, G_1) : G_0 \not\simeq G_1\}$$

.

Théorème 5.5. *Il y a un protocole HVZK pour GNI*

Preuve: On rappelle dans la figure 5.2 le protocole de preuve intéractive probabiliste vu la dernière fois.

Figure 5.2. Protocole NIP pour GNI

honnête	P		V	honnête
				$b \underset{\text{aléa}}{\longleftarrow} \{0,1\}$
				$\pi \underset{\text{aléa}}{\longleftarrow} \mathfrak{S}_n$
$G_0 \simeq H$?		$\overset{H}{\longleftarrow}$		$H = \pi(G_b)$
OUI :	$b' \leftarrow 0$			
NON :	$b' \leftarrow 1$	$\overset{b'}{\longrightarrow}$		
				accepte si $b = b'$
				rejette sinon

On a montré la dernière fois que pour $G_0 \not\simeq G_1$, le **(P)** honnête donne $b = b'$ avec probabilité 1 ; et pour $G_0 \simeq G_1$, pour tout P^*, $Pr[b = b'] \leq \frac{1}{2}$.

Voici un simulateur : $S(G_0, G_1)$ tire au sort b dans $\{0, 1\}$, et π dans $\mathfrak{S}_n$ retourne (G_0, G_1, b, π, b). □

C'est un exemple simple. Qu'est-ce qui n'est pas bon dans cette premiere definition ? Dans cet exemple, si **(V)** n'est pas honnete, il peut envoyer un graphe H obtenu d'une façon quelconque, et apprendre si H est isomorphe à G_0 ou pas.

Donc ce n'est pas un exemple de divulgation nulle (zéro knowledge) dans le sens intuitive. Il faut gérer le cas ou le verificateur n'est pas honnête. On modifie la definition precedente.

5.3 Divulgation nulle

Définition 5.6 (ZK : Zero knowledge ou divulgation nulle). *Un protocole* $\boldsymbol{P}$, $\boldsymbol{V}$ *pour un langage* $\mathcal{L}$ *est ZK si pour tout* ***(V*)*** *efficace et probabiliste, il existe* S *un "simulateur" efficace probabiliste tel que*

$$\forall x \in \mathcal{L}, S(x) \text{ et } View_{P,V^*}(x) \text{ sont statistiquement indistinguables}$$

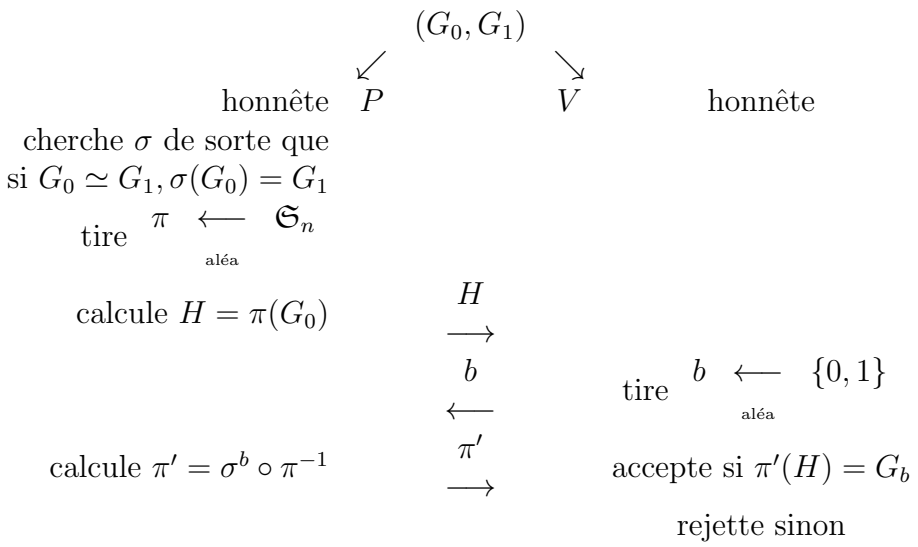
Considérons maintenant l'exemple du langage des graphes isomorphes

$$GI = \overline{GNI} = \{(G_0, G_1) : G_0 \backsimeq G_1\}$$

Théorème 5.7. *GI possede un protocole de preuve interactive probabiliste qui est ZK.*

Preuve: Dans la figure 5.3 on énonce le protocole pour un prouveur et un vérificateur honnêtes :

Figure 5.3. Protocole IP pour prouveur et vérificateurs honnêtes



Montrons que c'est bien un protocole "IP".

Soit $(G_0, G_1) \in GI$. Comme le tout-puissant **(P)** a correctement trouvé une permutation σ qui réalise l'isomorphisme entre G_0 et G_1 alors **(V)** accepte pour toutes les instances de π et b.

Soit $(G_0, G_1) \notin GI$, et soit P^* un prouveur quelconque, pas forcement honnête.

$$\begin{aligned} Pr[\langle P^*, V\rangle(G_0, G_1) \text{ accepte }] &= Pr[b=0]Pr[\pi'(H) = G_0] + Pr[b=1]Pr[\pi'(H) = G_1] \\ &= \frac{1}{2}(Pr[\pi'(H) = G_0] + Pr[\pi'(H) = G_1]) \\ &\ll \frac{1}{2} \end{aligned}$$

C'est parce que quel que soit le graphe H qu'envoie **P***, H est isomorphe au plus à un d'entre G_0 et G_1, alors **P*** ne peut répondre correctement que à une seule valeur de b, et donc il ne peut réussir que avec probabilité $1/2$.

Montrons maintenant que ce protocole est ZK. Il faut montrer

$$\forall V^*, \exists S \forall (G_0, G_1) \in GI : S(G_0, G_1) \text{ indistinguible de } VIEW_{P,V^*}(G_0, G_1)$$

Voici le simulateur $S(G_0, G_1)$:

- tirer au sort $\omega \leftarrow \{0,1\}^*$, aléa pour V^*
- Pour $i = 1$ à n (où n est la taille de l'entrée)
  - tirer au sort $b' \leftarrow \{0,1\}$, $\pi \leftarrow S_n$
  - calculer $b = V^*(\pi(G_b); \omega)$
  - si $b \neq b'$, répeter
  - si $b = b'$, retourner $(\omega, \pi(G_b), b, \pi^{-1})$
- Si $i = n$, échouer.

Si $(G_0, G_1) \in GI$, comme

$$\Pi(G_0) = \{\pi(G_0) | \pi \in S_n\} = \{\pi(G_1) | \pi \in S_n\} = \Pi(G_1),$$

alors $Pr[V^*(\pi(G_b), \omega) = b'] = \frac{1}{2}$, d'où

$$Pr[S(G_0, G_1) \text{ échoue }] \leq 2^{-n}$$

□

On va voir comment un autre exemple qui n'était pas encore ZK devient ZK.

Théorème 5.8. *Le langage GNI possède un protocole de preuve interactive aléatoire qui est ZK.*

Preuve: Considérons le protocole de la figure 5.4 pour le langage GNI.

Dans la description de ce protocole on a utilisé le symbole $\circ$ pour signifier la concatenation de deux graphes : il en résulte un graphe non-connexe à $2n$ sommets où le premier graphe se trouve sur les n premiers sommets et le deuxième graphe se trouve sur les n prochains sommets.

Si $(G_0, G_1) \in GNI$, le prouveur honnête envoie comme dernière réponse $b' = b$. en effet :
- Si $a_i = 0$, $\tau(J_i) = \rho_i^{-1}\rho_i(G) = G$
- Si $a_i = 1$, $\tau(J_i) = \pi \circ \rho_i^{-1} \circ \rho_i(G) = \pi(G) = H$

Si $(G_0, G_1) \notin GNI$, on veut borner pour tout P^* la probabilité $Pr[\langle P^*, V\rangle(G_0, G_1) \text{ accepte}]$

Intuitivement, comme on utilise un bout du protocole ZK pour GI, il ne peut pas augmenter l'information réçue par le prouveur donc ne peut pas permettre au prouveur de tricher.

Soit **P',V'** le protocole d'origine pour GNI (protocole sans vérification de $H \simeq G$). Pour tout P^* pour protocole avec vérification, il existe P^{**} pour protocole d'origine tel que

$$Pr[< P^*, V > (G_0, G_1) \text{ accepte}] \leq Pr[< P^{**}, V' > (G_0, G_1) \text{ accepte}] + \text{qqch de négligeable}$$

	P^{**}		V
pour $i = 1$ à n		H $\longleftarrow$	
execute simulateur GNI avec P^* comme **V**érificateur, obtient J_i, a_i, τ_i			
$b' = P^*(H, (J_i, a_i, \tau_i)_{i=1}^n)$		b' $\longrightarrow$	

Figure 5.4. Protocole ZK pour GNI

	(G_0, G_1)	
honnête P	$\swarrow$	$\searrow$ V honnête
		$G = G_0 \circ G_1$
		tire $b \xleftarrow[\text{aléa}]{} \{0,1\}$
		tire $\pi_0 \xleftarrow[\text{aléa}]{} \mathfrak{S}_n$
trouve b' tel que		tire $\pi_1 \xleftarrow[\text{aléa}]{} \mathfrak{S}_n$
$H = H_{b'} \circ H_{1-b'}$	$\xleftarrow{H}$	$H = \pi_0(G_b) \circ \pi_1(G_{1-b})$
$H_{b'} \simeq G_{b'}$		pour $i = 1$ à n
$H_{1-b'} \simeq G_{1-b'}$		tire $\rho_i \xleftarrow[\text{aléa}]{} \mathfrak{S}_{2n}$
	$\xleftarrow{J_i}$	calcule $J_i = \rho_i(G)$
tire $a_i \xleftarrow[\text{aléa}]{} \{0,1\}$	$\xrightarrow{a_i}$	
vérifie que	$\xleftarrow{\tau}$	calcule $\tau = \pi^{a_i} \circ \rho_i$
si $a_i = 0, \tau(J_i) = G$		
si $a_i = 1, \tau(J_i) = H$		
	$\xrightarrow{b'}$	accepte si $b = b'$
		rejette sinon

Convaincons nous de cette affirmation $|Pr[< P^*, V > (G_0, G_1) accepte] - Pr[< P^{**}, V' > accepte]| \leq \Delta$(interaction réelle, $S(G, H) \leq$ négligeable ce qui montre l'inégalité desirée.

$$Pr[< P^*, V > (G_0, G_1) accepte] \leq 1/2 + negligeable$$

On a utilisé que pour toute fonction f,

$$\Delta(f(X), f(Y)) \leq \Delta(X, Y)$$

Maintenant, il reste à vérifier la condition ZK :

$$\forall V^* \exists S, \forall (G_0, G_1) \in GNI, S(G_0, G_1) indistinguible de VIEW_{P,V^*}(G_0, G_1)$$

Voici le simulateur :
$S(G_0, G_1)$:
– $\omega \leftarrow \{0,1\}^*$ aléa pour **V***

- $H = V^*(G_0, G_1, 1; \omega)$
- pour $i = 1$ à n
  - $J_i = V^*(G_0, G_1,$ i-eme réponse au premier message, $a_1, .., a_{i-1}, \omega)$
  - tire $a_i \underset{\text{aléa}}{\longleftarrow} \{0,1\}$
  - $\tau_i^{a_i} = V^*(G_0, G_1,$ i-ème réponse au second message, $a_1, ..., a_i, \omega)$
  - Si
    - $a_i = 0$ alors vérifier $\tau_i^{a_i}(J_i) = G$
    - $a_i = 1$ alors vérifier $\tau_i^{a_i}(J_i) = H$
    - si ce n'est pas vérifié échouer
  - calculer $\tau_i^{1-a_i} = V^*(G_0, G_1, iemereponse, secondmessage, a_1, .., a_{i-1}, 1 - a_i, \omega)$
  - Si
    - $1 - a_i = 0$ alors vérifier $\tau_i^{1-a_i}(J_i) = G$
    - $1 - a_i = 1$ alors vérifier $\tau_i^{1-a_i}(J_i) = H$
    - si c'est vérifié, determiner b' : calculer $\pi = \tau_i^1 \circ (\tau_i^0)^{-1}$ (si π echange $1 \cdots n$ avec $n+1 \cdots 2n$ alors $b' = 1$, si π est dans $\mathfrak{S}_n \circ \mathfrak{S}_n$ alors $b' = 0$)
- si $i = n$ et b' n'est pas determiné, échouer
- sinon retourner $(G_0, G_1, \omega, (J_i, a_i, \tau_i), b')$

Lemme 5.9. $Pr[S(G_0, G_1)abort] \leq Pr[< P, V^* > (G_0, G_1)abort] + 2^{-n}$

Preuve:

$$\begin{aligned} Pr[S(G_0, G_1) \text{ échoue}] &= Pr[S(G_0, G_1) \text{ échoue quand } V^* \text{ donne mauvais } \tau_i^{a_i}] \\ &\quad + Pr[S(G_0, G_1) \text{ échoue quand } V^* \text{ donne bonnes réponses}] \\ &\leq Pr[\langle P, V^* \rangle (G_0, G_1) \text{ échoue}] + 2^{-n} \end{aligned}$$

Si V^* donne les bonnes réponses mais $S(G_0, G_1)$ échoue alors $\forall i = 1, \ldots, n$ $\exists c_i \in 0, 1$ tel que $V^*(G_0, G_1, \omega, c_1, \ldots, c_i) = \tau_i^{c_i}$ vérifie mais $V^*(G_0, G_1, \omega, c_1, \ldots, c_{i-1}, 1 - c_i) = \tau_i^{1-c_i}$ ne vérifie pas. Donc

$$Pr[S(G_0, G_1) \text{ échoue quand } V^* \text{ donne bonnes réponses}] \leq 2^{-n}.$$

□

Il reste à montrer que le point de vue du vérificateur et le simulateur sont statistiquement indistinguibles

$$VIEW_{P,V^*}(G_0, G_1) = (G_0, G_1, \omega, a_1, \ldots, a_n, b')$$

$\forall \omega, a_1, \ldots, a_n, b'$ dans le support de $S(G_0, G_1)$

$$\begin{aligned} Pr[VIEW_{P,V^*}(G_0, G_1) = (G_0, G_1, \omega, a_1, \ldots, a_n, b')] &= 2^{-t} 2^{-n} \\ &= Pr[S(G_0, G_1) = (G_0, G_1, \omega, a_1, \ldots, a_n, b')] \end{aligned}$$

ce qui permet de conclure. □

Dans ce cours, nous n'avons considéré que les preuves qui sont statistiquement zéro knowledge (SZK). Il existe aussi la notion de zéro knowledge calculatoire, où on demande que la sortie du simulateur et la vraie interaction soient indistinguables par les distingueurs efficaces (au lieu des distingueurs quelconques, comme dans la indistinguibilité statistique). Cette relaxation permet de montrer que tout langage qui a une preuve interactive a aussi une preuve interactive zéro knowledge