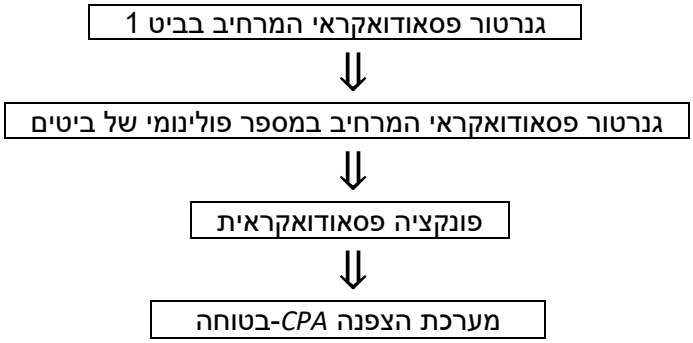
קריפטוגרפיה 2

הרצאה 7: פונקציות חד-כיווניות

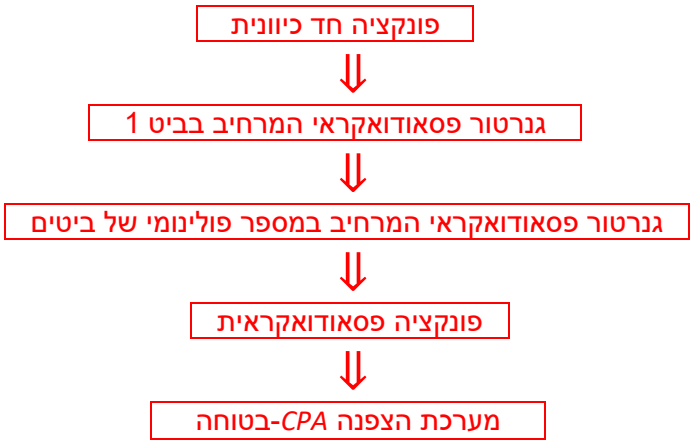
Textbook: Katz and Lindell. Introduction to Modern Cryptography.

מרצה: אורי שטמר

עד היום ראינו:



היינו רוצים לבנות מערכת הצפנה בטוחה על סמך ההנחה המינימאלית האפשרית. עד היום ראינו איך בונים מערכת הצפנה בטוחה מתוך יצרן פסאודואקראי. היום נגדיר פרימיטיב נוסף – פונקציה חד כיוונית – ונראה שממנה אפשר לבנות יצרן פסאודואקראי. כלומר נקבל:



<u>הגדרה:</u> נאמר כי פונקציה $f:\{0,1\}^* \to \{0,1\}^*$ היא <u>חד-כיוונית</u> אם:

(1) קיים אלגוריתם פולינומי דטרמניסטי שבהינתן x מחשב את $f(x)$

(2) קשה להפוך את f. כלומר, כל יריב פולינומי אקראי (אחיד או לא-אחיד) לא יכול להפוך את f בהסתברות לא זניחה. כלומר, לכל יריב פולינומי אקראי $\mathcal{A}$ קיימת פונקציה זניחה $negl$ כך שההסתברות ש- $\mathcal{A}$ מנצח במשחק הבא היא לכל היותר $negl(n)$.

<u>משחק $Invert(1^n)$:</u>

א. המאתגר בוחר $x \in \{0,1\}^n$ בהתפלגו אחידה ומחשב $y = f(x)$.

ב. היריב $\mathcal{A}$ מקבל $y, 1^n$ ומחזיר פלט x'. היריב מנצח אם $f(x') = y$.

<u>שאלה:</u> למה שלא נדרוש שהיריב צריך להחזיר את x?

<u>תשובה:</u> כי אז הפונקציה $f(x_1,x_2)=x_1$ תהייה חד-כיוונית אבל זאת לא פונקציה שימושית...

<u>שאלה:</u> תהי f פונקציה שניתנת לחישוב בזמן פולינומי אשר <u>אינה</u> פונקציה חד-כיוונית. מה זאת אומרת?

<u>תשובה:</u> קיים יריב שמנצח במשחק $Invert(1^n)$ בהסתברות $\epsilon(n)$, עבור $\epsilon(n)$ פונקציה <u>לא</u> זניחה.
תזכורת: פונקציה $\epsilon(n)$ היא זניחה אם לכל פולינום חיובי $p(n)$ קיים n_0 כך שלכל $n>n_0$ מתקיים $\epsilon(n)\leq\frac{1}{p(n)}$. פונקציה $\epsilon(n)$ <u>איננה</u> זניחה אם קיים פולינום חיובי $p(n)$ כך שלאינסוף ערכי n מתקיים $\epsilon(n)>\frac{1}{p(n)}$.

<u>לאן אנחנו הולכים?</u>
נניח שיש לנו פונקציה חד-כיוונית f. אנחנו רוצים לבנות PRG שמרחיב בביט אחד.

<u>אינטואיציה שגויה:</u>
מכיוון ש- f היא חד-כיוונית אזי מתוך $f(x)$ אי אפשר ללמוד את x בשלמותו. תקווה שגוייה: אולי זה אומר שמתוך $f(x)$ אי אפשר ללמוד "כלום" על x, כלומר x עדיין "נראה אקראי" גם בהינתן $f(x)$ ולכן נוכל לבנות PRG באופן הבא:

$$G(\underbrace{x}_{\text{ביטים } n})=\underbrace{f(x),x_1}_{\text{ביטים } n+1}$$

הבעייה כאן היא שזה שאי אפשר לחשב את x בשלמותו <u>לא</u> אומר שאי אפשר לחשב הרבה ביטים של x.

לדוגמה, אם f חד-כיוונית אז גם $f'(x_1,x_2)=\big(x_1,f(x_2)\big)$ היא חד-כיוונית (למה?).

<u>נסיון שגוי נוסף:</u>
אולי אם f היא חד-כיוונית אז קשה מתוך $f(x)$ לחשב את $\bigoplus_{i=1}^{n}x_i$ (לכאורה דורש לחשב כל ביט).

<u>דוגמה נגדית:</u>
אם f חד-כיוונית אז גם $f'(x)=f(x),\bigoplus_{i=1}^{n}x_i$ חד-כיוונית.

הוכחה ש- f' חד-כיוונית: אם מתוך $f'(x)$ אפשר לחשב את x (או איבר אחר ב- $\big(f'(x)\big)^{\prime(-1)}$ אז נבנה את היריב הבא ל- f: בהינתן $f(x)$ נריץ את היריב על $f(x),0$ ועל $f(x),1$ ונבדוק אם קיבלנו איבר ב f^{-1}. מכיוון שבאחד המקרים נתנו את $\bigoplus_{i=1}^{n}x_i$ אז נצליח להפוך את f בהסתברות לא זניחה.

נראה שהכללה של רעיון ה XOR כן תעבוד. אבל לפני שנמשיך, נציג עוד הגדרה שתקל לנו על החיים. אמרנו שנראה שמתוך פונקציה חד-כיוונית אפשר לבנות PRG, אבל כדי להקל על עצמנו נתמקד רק בסוג <u>מסויים</u> של פונקציות חד-כיווניות ונראה שהסוג המסויים הזה גורר PRG.

<u>הגדרה:</u>
פונקציה חד-כיוונית f היא <u>פרמוטציה חד-כיוונית</u> אם לכל x מתקיים $|f(x)|=|x|$ ולכל n הפונקציה
$$f_n(x)=f(x),\qquad f_n:\{0,1\}^n\to\{0,1\}^n$$
היא פרמוטציה.

<u>הגדרה:</u>
פונקציה $hc:\{0,1\}^* \to \{0,1\}$ היא <u>ביט קשה</u> (hard-core) עבור פונקציה f אם:
1. hc ניתנת לחישוב בזמן פולינומי
2. לכל יריב יעיל $\mathcal{A}$ קיימת פונקציה זניחה $negl$ כך שלכל n מתקיים

$$\Pr_{\substack{x\in\{0,1\}^n \\ \mathcal{A}}}\left[\mathcal{A}(1^n, f(x)) = hc(x)\right] \leq \frac{1}{2} + negl(n)$$

כאשר ההסתברות היא מעל הגרלת x בהתפלגות אחידה מ- $\{0,1\}^n$ ומעל האקראיות של $\mathcal{A}$.

<u>דוגמה לא מעניינת:</u> $f(x_1 \dots x_n) = x_1 \dots x_{n-1}$ ו- $hc(x_1 \dots x_n) = x_n$.
פונקציה כזאת לא תעור לנו לבנות PRG...

אנחנו נתעניין בביטים קשים של <u>פרמוטציה</u> חד-כיוונית.
המסר הוא שפונקציה חד-כיוונית יכולה לאבד מידע על הקלט. פרמוטציה לעומת זאת לא מאבדת מידע על הקלט ולמרות זאת אין אלגוריתם יעיל שהופך פרמוטציה חד-כיוונית.

<u>דוגמאות לביט קשה:</u>

((1)) עבור הלוג הדיסקרטי $f_{p,g}(x) = g^x \bmod p$ נגדיר

$$hc: \mathbb{Z}_p^* \to \{0,1\}$$
$$hc(x) = \text{msb}(x)$$

* כפי שראינו בקורס הקודם, ה lsb <u>איננו</u> ביט קשה כי אפשר לחשב אותו ע"י בדיקה אם g^x הוא שארית ריבועית או לא $\left(a^{\frac{p-1}{2}} \equiv 1 \pmod p \iff a \in QR_p\right)$.
* עבור p-ים מסויימים נקבל אפילו <u>שכמה</u> מהביטים ה lsb הם לא קשים
* אבל בכל מקרה, החל מנקודה מסויימת והלאה לכיוון ה msb כל ביט יהיה קשה

((2)) עבור *RSA* כלומר עבור $f_{N,e}(x) = x^e \bmod N$ כאשר $N = p \cdot q$ מכפלת ראשוניים, נגדיר

$$hc: \mathbb{Z}_N \to \{0,1\}$$
$$hc(x) = \text{lsb}(x) = x \bmod 2$$

- כאן למעשה כל ביט של x יהיה קשה

לא ידוע אם לכל פונקציה (פרמוטציה) חד-כיוונית יש ביט קשה. ידועה תוצאה כמעט כזאת: לכל פונקציה (פרמוטציה) חד-כיוונית f קיימת פונקציה (פרמוטציה) חד-כיוונית g שיש לה ביט קשה. זה מספיק בשביל לבנות גנרטור פסאודואקראי מתוך פונקציה (פרמוטציה) חד-כיוונית.

<u>משפט (Oded Goldreich and Leonid Levin, 1989):</u> תהי f פונקציה חד-כיוונית ונגדיר $g(x,r) = (f(x), r)$ עבור $|x| = |r|$, ונגדיר $GL(x,r) = \oplus x_i \cdot r_i$ (מכפלה פנימית מודולו 2). אזי GL היא ביט קשה עבור g.

כדי להוכיח את המשפט הזה אנחנו צריכים להראות שלא קיים אלגוריתם יעיל שבהינתן $g(x,r)$ מצליח לנחש את $GL(x,r)$ בהסתברות $> \frac{1}{2} + negl(n)$. לפני שנראה את ההוכחה המלאה, נראה 2 חימומים:

<u>חימום 1</u>: נראה שלא קיים אלגוריתם שמנחש את $GL(x,r)$ בהסתברות 1

חימום 2: נראה שלא קיים אלגוריתם שמנחש את $GL(x,r)$ בהסתברות $\gg \frac{3}{4}$

<u>חימום 1: טענה:</u> יהיו f,g,GL כמו במשפט. אם קיים אלגוריתם פולינומי $\mathcal{A}$ כך ש- $\mathcal{A}(f(x),r)=GL(x,r)$ לכל n ולכל $x,r\in\{0,1\}^n$ אזי קיים אלגוריתם פולינומי $\mathcal{B}$ כך ש- $\mathcal{B}\big(f(x)\big)=x$ לכל n ולכל $x\in\{0,1\}^n$.

<u>הוכחה:</u>
נשים לב ש- $GL(x,e_i)=x_i$ לפי הגדרת GL, כאשר $e_i=(0,0,\dots,0,1,0,\dots,0)$ הוא ווקטור היחידה ה-i-אי.

<u>אלגוריתם $\mathcal{B}$ על קלט y:</u>
(1) לכל $1\le i\le n$ הרץ $\mathcal{A}$ על (y,e_i) וקבל x_i
(2) החזר $x_1,\dots,x_n$

מסקנה: אם f חד-כיוונית אזי לא קיים $\mathcal{A}$ כנ"ל.

<u>חימום 2: טענה:</u> יהיו f,g,GL כמו במשפט. אם קיים אלגוריתם פולינומי אקראי $\mathcal{A}$ ופולינום p כך שלאינסוף ערכי n מתקיים

$$\Pr_{\substack{x\in\{0,1\}^n\\ r\in\{0,1\}^n\\ \mathcal{A}}}[\mathcal{A}(f(x),r)=GL(x,r)]\ge\frac{3}{4}+p(n)$$

אזי קיים אלגוריתם פולינומי אקראי $\mathcal{B}$ כך שלאותם ערכי n מתקיים:

$$\Pr_{x,\mathcal{B}}\big[\mathcal{B}\big(f(x)\big)\in f^{-1}\big(f(x)\big)\big]\ge\frac{1}{4\cdot p(n)}$$

(ולכן אם f חד-כיוונית אז לא קיים $\mathcal{B}$ כזה ולכן לא קיים $\mathcal{A}$ כנ"ל)

<u>דיון:</u> מה הבעייה עכשיו? עכשיו $\mathcal{A}$ לא תמיד צודק. בפרט, יתכנו ערכים של x,r עליהם $\mathcal{A}$ תמיד טועה. למשל יתכן שעל r-ים מהצורה $r=e_i$ תמיד $\mathcal{A}$ טועה (ההסברות ש- r אקראי יהיה וקטור יחידה היא זניחה ולכן ל- $\mathcal{A}$ מותר תמיד לטעות על r-ים כאלה...)

<u>נסיון להתגבר על הבעייה:</u> במקום להריץ את $\mathcal{A}$ על $\mathcal{A}\big(f(x),e_j\big)$ כדי ללמוד את x_j, נגריל r ונריץ את $\mathcal{A}$ פעמיים על $\mathcal{A}\big(f(x),r\oplus e_j\big)$ ועל $\mathcal{A}(f(x),r)$. בתקווה נקבל את $GL\big(x,r\oplus e_j\big)$ ואת $GL(x,r)$ ומתוכם נוכל לחשב:

$$GL(x,r)\oplus GL\big(x,r\oplus e_j\big)=\left(\bigoplus_{i=1}^{n}x_i\cdot r_i\right)\oplus\left(\bigoplus_{i=1}^{n}x_i\cdot(r_i\oplus e_{ji})\right)$$
$$=\underbrace{\left(\bigoplus_{i=1}^{n}x_i\cdot r_i\right)\oplus\left(\bigoplus_{i=1}^{n}x_i\cdot r_i\right)}_{\text{מצטמצמים}}\oplus\underbrace{\left(\bigoplus_{i=1}^{n}x_i\cdot e_{ji}\right)}_{=x_j}=x_j$$

<u>למה זה טוב לנו?</u>
עכשיו אנחנו מריצים את $\mathcal{A}$ על $\mathcal{A}(f(x),r)$ עבור x,r אקראיים כמו ש- $\mathcal{A}$ מצפה, ולכן הוא מצליח לחשב את $GL(x,r)$ בהסתברות $>3/4$. כנ"ל, $\mathcal{A}\big(f(x),r\oplus e_j\big)$ מצליח לחשב את $GL\big(x,r\oplus e_j\big)$ בהסתברות $>3/4$ (אם r מתפלג אחיד אזי גם $r\oplus e_j$ מתפלג אחיד). לכן, מחסם האיחוד, בהסתברות קצת יותר מ- $1/2$ נצליח לחשב את x_j.

<u>למה זה עדיין לא סוף הסיפור?</u>

הסתברות קצת יותר מחצי עבור כל קוארדינטה בודדת זה לא מספיק. נרצה להגביר את הביטחון שלנו ע"י כך שנחזור על התהליך הזה הרבה פעמים עם r-ים אקראיים וניקח את החלטת הרוב. כלומר היינו רוצים להגריל הרבה r-ים אקראיים, עם כל אחד מהם לחשב ניחוש $x_j = \mathcal{A}(f(x),r) \oplus \mathcal{A}(f(x),r \oplus e_j)$ ואז לקחת את החלטת הרוב.

אבל אנחנו צריכים להיזהר. מה שמובטח לנו זה ש- $\mathcal{A}$ מצליח (בהסתברות $> 3/4$) כש- x,r נבחרים באקראי (שניהם). המצב כאן הוא שונה. x נקבע פעם אחת ואנחנו רוצים לחשב את $\mathcal{A}$ על $f(x),r$ עם אותו x ועם הרבה r-ים שונים. האם מובטח לנו ש- $\mathcal{A}$ יצליח בהסתברות גבוהה במקרה כזה?

הסכנה היא שייתכן שיש מחרוזות x שאם נקבע אותם ואז נבחר את r באקראי אז $\mathcal{A}$ לא מצליח בהסתברות טובה. כלומר יתכן שרק כשאנחנו ממצעים על פני (x,r) אקראיים אזי $\mathcal{A}$ מצליח בהסתברות $> 3/4$ אבל יתכן שיש x-ים "טובים" שעליהם $\mathcal{A}$ מצליח (בהסתברות טובה) להרבה r-ים ושיש x-ים "רעים" שעליהם $\mathcal{A}$ מצליח (בהסתברות טובה) רק למעט r-ים.

מה שנראה עכשיו זה שעבור $\mathcal{A}$ כמו בטענה שלנו חייבים להיות הרבה x-ים "טובים".

<u>למה:</u> יהי n כך ש-

$$\Pr_{\mathcal{A},x,r}[\mathcal{A}(f(x),r) = GL(x,r)] \geq \frac{3}{4} + \frac{1}{p(n)}$$

אזי קיימת $s_n \subseteq \{0,1\}^n$ שגודלה $|S_n| \geq \frac{2^n}{2 \cdot p(n)}$ כל שלכל $x \in S_n$ מתקיים

$$\Pr_{\mathcal{A},r}[\mathcal{A}(f(x),r) = GL(x,r)] \geq \frac{3}{4} + \frac{1}{2 \cdot p(n)}$$

הלמה הזאת אומרת לנו שחייבים להיות הרבה ערכים של x שאחרי שנקבע אותם נוכל להרגיל הרבה ערכי r לאותו $f(x)$ ו- $\mathcal{A}$ יצליח (בהסתברות $> 3/4$) לחשב את $GL(x,r)$ המתאים.

<u>הוכחת הלמה:</u> נגדיר

$$S_n = \left\{ x \in \{0,1\}^n \quad : \quad \Pr_r[\mathcal{A}(f(x),r) = GL(x,r)] \geq \frac{3}{4} + \frac{1}{2 \cdot p(n)} \right\}$$

נחשב:

$$\frac{3}{4} + \frac{1}{p(n)} \leq \Pr_{\mathcal{A},x,r}[\mathcal{A}(f(x),r) = GL(x,r)] =$$

$$= \Pr[x \in S_n] \cdot \underbrace{\Pr[\mathcal{A}(f(x),r) = GL(x,r) | x \in S_n]}_{\leq 1} + \underbrace{\Pr[x \notin S_n]}_{\leq 1} \cdot \underbrace{\Pr[\mathcal{A}(f(x),r) = GL(x,r) | x \notin S_n]}_{\leq \frac{3}{4} + \frac{1}{2 \cdot p(n)} \quad (\text{לפי הגדרת } S_n)}$$

$$\leq \Pr[x \in S_n] + \frac{3}{4} + \frac{1}{2 \cdot p(n)}$$

ולכן

$$\frac{|S_n|}{2^n} \underbrace{=}_{\substack{\text{כי בוחרים} \\ \text{את } x \text{ באקראי}}} \Pr[x \in S_n] \geq \frac{1}{2 \cdot p(n)}$$

כלומר

$$|S_n| \geq \frac{2^n}{2 \cdot p(n)}$$

מה שמוכיח את הלמה.

<u>הוכחת הטענה של חימום 2:</u>

<u>אלגוריתם $\mathcal{B}$ להיפוך f:</u>

קלט: $y = f(x)$

- עבור $i = 1,2,\dots,n$
  - בחר $r \in \{0,1\}^n$ בהתפלגות אחידה וחשב ניחוש ל- x_i:
    $\tilde{x}_i = \mathcal{A}(f(x), r) \oplus \mathcal{A}(f(x), r \oplus e_i)$
  - חזור על התהליך "מספיק פעמים" וקבע
    $z_i =$ החלטת הרוב על פני הניחושים $\tilde{x}_i$
- הפלט: $z = z_1 z_2 \dots z_n$

<u>ניתוח:</u> עבור $x \in S_n$ מתקיים

$$\Pr_{\mathcal{A},r}[\mathcal{A}(f(x),r) \oplus \mathcal{A}(f(x), r \oplus e_i) \neq x_i]$$

$$\leq \Pr_{\mathcal{A},r}[\mathcal{A}(f(x),r) \neq GL(x,r) \quad OR \quad \mathcal{A}(f(x), r \oplus e_i) \neq GL(x, r \oplus e_i)]$$

$$\leq \underbrace{\Pr_{\mathcal{A},r}[\mathcal{A}(f(x),r) \neq GL(x,r)]}_{\leq \frac{1}{4} - \frac{1}{2 \cdot p(n)}} + \underbrace{\Pr_{\mathcal{A},r}[\mathcal{A}(f(x), r \oplus e_i) \neq GL(x, r \oplus e_i)]}_{\leq \frac{1}{4} - \frac{1}{2 \cdot p(n)}} \leq \frac{1}{2} - \frac{1}{p(n)}$$

מיד נראה שעם מספיק חזרות, ההסתברות ש- $z_i \neq x_i$ היא לכל היותר $\frac{1}{2n}$ ולכן (עבור $x \in S_n$) ההסתברות ש- $z \neq x$ היא

$$\Pr[z \neq x] = \Pr\left[\bigvee_{i=1}^{n} \{z_i \neq x_i\}\right] \leq \sum_{i=1}^{n} \Pr[z_i \neq x_i] \leq n \cdot \frac{1}{2n} = \frac{1}{2}$$

לכן הסתברות ההצלחה של $\mathcal{B}$ בהיפוך היא לפחות

$$\Pr[x \in S_n] \cdot \Pr[z = x | x \in S_n] \geq \frac{1}{2 \cdot p(n)} \cdot \frac{1}{2} = \frac{1}{4 \cdot p(n)}$$

בסתירה לכן ש- f חד-כיוונית.

אז כל מה שנותר לנו לעשות זה לחשב את מספר החזרות m. לצורך כך נשתמש במשפט הבא מהסתברות:

<u>חסם צ'רנוף:</u>

יהיו $v_1, \dots, v_m$ משתנים מקריים בלתי תלויים כך ש- $v_i \in \{0,1\}$ וכל שלכל i מתקיים $\Pr[v_i = 1] = \mathbb{E}[v_i] = \mu$.
אזי לכל $\epsilon > 0$ מתקיים:

$$\Pr\left[\left|\frac{\sum_{i=1}^{n} v_i}{m} - \mu\right| \geq \epsilon\right] \leq 2 \cdot e^{-2 \cdot \epsilon^2 \cdot m}$$

נגדיר משתנים מקריים $v_1, \dots, v_m$ כאשר $v_j = 1$ אם בהפעלה ה-j-ית מתקיים
$\mathcal{A}(f(x), r) \oplus \mathcal{A}(f(x), r \oplus e_i) = x_i$
ואחרת $v_j = 0$.

עבור $x \in S_n$ מתקיים $\mathbb{E}[v_j] \geq \frac{1}{2} + p(n)$. נחשב:

$$\Pr[z_i \neq x_i] = \Pr\left[\frac{\sum_{i=1}^{n} v_i}{m} \leq \frac{1}{2}\right] \leq \Pr\left[\left|\frac{\sum_{i=1}^{n} v_i}{m} - \frac{1}{2} - \frac{1}{p(n)}\right| \geq \frac{1}{p(n)}\right] \leq 2 \cdot e^{-\frac{2m}{p^2(n)}}$$

ולכן מספיק לקחת $m \approx p^2(n) \cdot \ln(n)$ כדי שההסתברות תהייה קטנה מ- $\frac{1}{2n}$.