


دانشکده‌ی علوم ریاضی

مقدمه‌ای پیشرفته بر رمزنگاری ۲ آبان ۱۳۹۱

جلسه‌ی ۱۲: مولد شبه‌تصادفی، سیستم امن تک‌پیامی

مدرس: شهرام خزائی *نگارنده: محمد قیاسی*

۱ یادآوری

تعریف ۱ *تابع $G : \{0,1\}^* \longrightarrow \{0,1\}^*$ یک مولد شبه‌تصادفی (PRG[1]) نامیده می‌شود اگر در شرایط زیر صدق کند:*

- *(محاسبه‌ی کارا) الگوریتم PPT ای وجود داشته باشد که G را محاسبه کند.*
- *(گسترش وردی) رشته‌های به طول n را به رشته‌های به طول $l(n)$ که ضریب گسترش[2] نامیده می‌شود ببرد که $l(n)$ یک چند جمله‌ای است و برای هر عدد طبیعی n داریم که : $l(n) > n$.*
- *$G(U_n)$ و $U_{l(n)}$ تمایزناپذیر محاسباتی باشند.*

تعریف ۲ *(جایگشت یک‌طرفه) تابع یک‌طرفهی $f : \{0,1\}^* \longrightarrow \{0,1\}^*$ یک جایگشت یک‌طرفه (OWP[3]) نامیده می‌شود اگر برای هر $n \geq 0$ تحدید f به $\{0,1\}^n$ یک جایگشت باشد.*

۲ ساخت مولد شبه‌تصادفی

در ابتدای این بخش خیلی سریع به چند قضیه‌ای که می‌خواهیم آنها را در طی این بخش بررسی کنیم، اشاره می‌کنیم و سپس در مورد آنها صحبت می‌کنیم.

قضیه ۱ *تابع یک‌طرفه وجود دارد اگر و فقط اگر مولد شبه‌تصادفی وجود داشته باشد.*

قضیه ۲ *اگر جایگشت یک‌طرفه وجود داشته باشد مولد شبه‌تصادفی هم وجود دارد.*

این قضیه به راحتی از قضیه‌های بعدی نتیجه می‌شود.

قضیه ۳ *اگر جایگشت یک‌طرفه وجود داشته باشد، مولد شبه‌تصادفی با ضریب گسترش $n+1$ وجود دارد.*

[1] psudo-random generator
[2] expansion factor
[3] one-way permutation

برای اثبات این قضیه سعی می‌کنیم که یک بیت به سمت راست خروجی جایگشت یک‌طرفه‌مان اضافه کنیم. این موضوع را به طور مفصل شرح خواهیم داد.

قضیه ۴ *اگر مولد شبه‌تصادفی با گسترش $n+1$ وجود داشته باشد به ازای هر چند جمله‌ای $l(\cdot)$ مولد شبه‌تصادفی با گسترش $l(n)$ وجود دارد.*

با تکرار کاری شبیه قضیه‌ی بالا، یعنی با افزودن بیت‌های پی‌درپی سعی می‌کنیم که طول خروجی را افزایش دهیم. این موضوع را نیز جلوتر به طور کامل‌تری شرح می‌دهیم.
در اینجا برای اثبات برخی قضیه‌های بالا نیاز داریم که یک مفهوم جدید را تعریف کنیم.

تعریف ۳ *یک تابع $h : \{0,1\}^* \longrightarrow \{0,1\}$ را یک تابع هاردکور برای $f(x)$ گوییم هرگاه h برای هر مقدار داده شده‌ی x به صورت کارا محاسبه پذیر باشد و هم چنین برای هر مهاجم چند جمله‌ای غیر یکنواخت $\mathcal{A}$ یک تابع ناچیز $\varepsilon(n)$ وجود داشته باشد که برای هر عدد طبیعی n داشته باشیم:*

$$\Pr\{x \longleftarrow \{0,1\}^n : \mathcal{A}(1^n, f(x)) = h(x)\} \leq \frac{1}{2} + \varepsilon(n)$$

برای مثال (تحت در فرض RSA) تابع $\mathrm{half}_N(x)$ که برابر ۱ است اگر و تنها اگر که $0 \leq x \leq \frac{N}{2}$ یک هاردکور برای RSAاست. حال با این تعریف سعی می‌کنیم قضیه‌ی ۳ را که در بالا شرح دادیم، بررسی کنیم. برای این کار همان قضیه را در قالب خاص تری بیان کرده و اثبات می‌کنیم.

قضیه ۵ *اگر f یک جایگشت یک‌طرفه باشد و h یک هاردکور برای آن باشد آنگاه $G(s) = f(s)||h(s)$ یک مولد شبه‌تصادفی است.*

برهان. فرض خلف می‌کنیم که یک مهاجم چند جمله‌ای غیر یکنواخت $\mathcal{A}$ و یک چند جمله‌ای $p(n)$ وجود دارد به طوری که برای نامتناهی تا مقدار طبیعی n یک i وجود دارد که $\mathcal{A}$، i امین بیت را به احتمال $\frac{1}{p(n)}$ پیش بینی می‌کند. ولی با توجه به اینکه n تا بیت اول $G(s)$ یک جایگشت از توزیع یکنواخت هستند تنها جای ممکن برای حدس زده شدن بیت آخر است که در نتیجه به دست می‌آید که:

$$\Pr\{\mathcal{A}(f(s)) = h(s)\} > \frac{1}{2} + \frac{1}{p(n)}$$

که این با این فرض که h یک هاردکور است در تناقض است. ■
حال با بسط PRG ای که در بالا ساختیم PRGهای با ضریب گسترش بزرگتر می‌سازیم.

لم ۶ *فرض کنید که $G : \{0,1\}^n \longrightarrow \{0,1\}^{n+1}$ یک PRG باشد. آنگاه برای هر چندجمله ای $l(n)$، مولد G' که به صورت زیر تعریف می‌شود شبه‌تصادفی است.*

$$G' : \{0,1\}^n \longrightarrow \{0,1\}^{l(n)}$$

را به صورت زیر تعریف می‌کنیم:

$$G'(s) = b_1 b_2 b_3 \cdots b_{l(n)}$$

که در آن $b_1, b_2, b_3, \cdots, b_{l(n)}$ *به طور بازگشتی زیر تعریف می‌شوند:*

$$X_0 = s$$

$$X_{i+1}||b_{i+1} \longleftarrow G(X_i)$$

یعنی، n *تا بیت اول* $G(X_i)$*،* X_{i+1} *است و بیت آخر آن* b_{i+1} *است. آنگاه* G' *یک* PRG *است.*

برهان. در ابتدا یک دنباله از مولدهای G^i را که

$$G^i : \{0,1\}^n \longrightarrow \{0,1\}^i$$

به طور بازگشتی به صورت زیر تعریف می‌کنیم:

$$x'||b \longleftarrow G(x)$$

$$G^0(x) = \varepsilon$$
$$G^i(x) = b||G^{i-1}(x')$$

که منظور از ε رشته‌ی تهی است. در ابتدا توجه می‌کنیم که جمله‌ی $l(n)$ ام دنباله‌ی فوق همان $G'(\cdot)$ است. حال فرض کنید که یک تمایزگر $\mathcal{D}$ و یک چند جمله‌ای $p(n)$ وجود دارند که $\mathcal{D}$ با احتمال بیش از معکوس $p(n)$ می‌تواند توزیع‌های $\{U_{l(n)}\}_n$ و $\{G'(U_n)\}_n$ را از هم تمایز بدهد. حال دنباله‌ی‌هایبرید

$$H_n^i = U_{l(n)-i}||G^i(U_n)$$

را برای $i = 1, 2, ..., l(n)$ تعریف می‌کنیم. در ابتدا توجه کنید که

$$H_n^0 = U_i, H_n^{i+1} = G^{l(n)}(U_n)$$

بنابر این $\mathcal{D}$ جملات H_n^0 و $H_n^{l(n)}$ را با احتمال معکوس چندجمله ای از هم تشخیص می‌دهد. با استفاده از لم هایبرید نتیجه می‌گیریم که برای یک i نامتناهی تا n وجود دارد که $\mathcal{D}$ توزیع‌های H_n^i و H_n^{i+1} را با احتمال $\frac{1}{l(n)p(n)}$ از هم تمایز می‌دهد. حال توجه کنید که

$$H_n^i = U_{l(n)-i}||G^i(U_n) = U_{l(n)-i-1}||U_1||G^i(U_n)$$

$$H_n^{i+1} = U_{l(m)-i-1}||G^{i+1}(U_n) = U_{l(n)-i-1}||b||G^i(x); \text{where } x||b \longleftarrow G(U_n)$$

و الگوریتم $M(y)$ را که nuPPT است و به صورت زیر تعریف می‌شود را در نظر بگیرید:

$$b_{prev} \longleftarrow U_{l(n)-i-1}$$

$$b \longleftarrow y_1$$
$$b_{next} \longleftarrow G^i(y_2...y_{n+1})$$
$$Output : b_{prev}||b||b_{next}$$

در مورد الگوریتم تعریف شده‌ی فوق می‌توان دید که : $M(U_{n+1}) = H_n^i$ و $M(G(U_n)) = H_n^{i+1}$. اما طبق فرض اولیه مسئله مبنی بر مولد شبه‌تصادفی بودن G، $\{G(U_n)\}_n$ و $\{U_{n+1}\}_n$ تمایزناپذیر محاسباتی هستند و در نتیجه تحت اثر الگوریتم کارای M تمایزناپذیر محاسباتی باقی می‌مانند که این به این معناست که H_{nn}^i و ${H_n^{i+1}}_n$ تمایزناپذیر محاسباتی اند و این یک تناقض است. ■

حال قضیه ای دیگر را معرفی می‌کنیم که در واقع نتیجه ای از قضیه‌ی قبل است:

قضیه ۷ *فرض کنید که f یک جایگشت یک‌طرفه باشد و h یک هاردکر برای آن. آنگاه*

$$G(x) = h(x)||h(f(x))||h(f^2(x))||...||h(f^{l(n)}(x))$$

یک مولد شبه‌تصادفی است.

برهان. می‌دانیم که اگر $G'(x) = f(x)||h(x)$ یک مولد شبه‌تصادفی است. حال اگر قضیه‌ی پیشین را که در مورد گسترش یک مولد شبه‌تصادفی بود را در مورد G' اعمال کنیم به راحتی حکم اثبات می‌شود. ■

به عنوان مثال با در نظر گرفتن فرض لگاریتم گسسته مولد زیر شبه‌تصادفی است:

$$G(x) = \text{half}_{p-1}(x)||\text{half}_{p-1}(g^x)||...$$

به این مولد Blum-Micali می‌گوییم. یا به عنوان یک مثال دیگر با در نظر گرفتن فرض RSAمولد زیر شبه‌تصادفی است:

$$G(x) = \text{LSB}(x)||\text{LSB}(x^e)||...$$

که $N = pq$و $x \in z_N^*$ و $\text{LSB}(\cdot)$ هارکور RSAاست.

۳ سامانه‌ی رمز امن

حال که می‌توانیم یک مولد شبه‌تصادفی بسازیم می‌خواهیم به کمک آن یک سیستم رمز بسازیم.

تعریف ۴ *سیستم رمز متقارن $(\text{Gen}, \text{Enc}, \text{Dec})$ امن تک‌پیامی‌نامیده می‌شود اگر:*

$$\forall n \in N, \forall m_1, m_2 \in \{0,1\}^n$$

دو متغیر تصادفی روبرو تمایزناپذیر محاسباتی باشند:

$$\{k \longleftarrow \text{Gen}(1^n) : \text{Enc}(m_0)\}, \{k \longleftarrow \text{Gen}(1^n) : \text{Enc}(m_1)\}$$

حال یک سیستم رمز می‌سازیم که طول کلیدش از طول پیامش کمتر و طبق تعریف فوق امن تک‌پیامی محسوب می شود.

قضیه ۸ *سیستم رمز زیر امن تک‌پیامی است.*

$$\text{Gen}(1^n) : k \longleftarrow U_{\frac{n}{2}}$$
$$\text{Enc}_k(m) = m \oplus G(k)$$
$$\text{Dec}_k(m) = m \oplus G(k)$$

که منظور از G در آن بیک مولد شبه‌تصادفی روی فضای رشته‌هاست.

برهان. برای اثبات اینکه این سیستم امن تک‌پیامی است باید بگوییم که به ازای هر $m_۱, m_۲$ ای دو توزیع زیر تمایزناپذیر محاسباتی اند. برای این کار از لم هایبرید استفاده می‌کنیم و ۴ تا جمله‌ی زیر را به عنوان دنباله مان تعریف می‌کنیم:

$$H^۱ = k \longleftarrow \mathsf{Gen}(۱^n) : m_۱ \oplus G(k)$$

$$H^۲ = u \longleftarrow U_n : m_۱ \oplus u$$

$$H^۳ = u \longleftarrow U_n : m_۲ \oplus u$$

$$H^۴ = k \longleftarrow \mathsf{Gen}(۱^n) : m_۲ \oplus G(k)$$

حال اگر جملات اول و آخر تمایز پذیر محاسباتی باشند طبق لم هایبرید حداقل دو جمله‌ی متوالی وجود دارند که تمایزناپذیر محاسباتی اند. برای این بررسی این مورد حالت بندی می‌کنیم و نشان می‌دهیم که در هر یک از حالات، جملات متوالی تمایز پذیر محاسباتی نیستند. جملات ۱ و ۲ که تمایز پذیر محاسباتی نیستند زیرا اگر یک تمایزگری بتواند آنها را از هم تمایز دهد می‌توانیم از آن برای تمایز دادن $G(k)$ و $U(n)$ استفاده کنیم. به این صورت که اگر دو تا نمونه که یکی از $G(k)$ و دیگری از $U(n)$ آمده اند را در نظر بگیریم برای تمایز دادن اینها از هم کافی است هر دو را با $m_۱$ جمع XOR کنیم و خروجی‌ها رار به تمایزگرمان بدهیم آنگاه او می‌تواند با احتمال قابل توجهی برای نمونه‌های مختلف اینها را از هم تشخیص بدهد. تمایزناپذیری سومی و چهارمی هم به همین صورت است و در مورد دومی و سومی هم می‌توان دیدی که هر دو توزیع داده شده همان توزیع یکنواخت هستند. ■