به نام خدا

دانشکده‌ی علوم ریاضی

مقدمه‌ای بر رمزنگاری ۶ اسفند ۱۳۹۱

جلسه‌ی ۶: مولد شبه‌تصادفی و سیستم رمز دنباله‌ای

مدرس: دکتر شهرام خزائی *نگارنده: مریم باروتی*

در جلسات قبل سیستم رمز OTP را معرفی کردیم که برای رمز کردن پیام m، کلید تصادفی k را که طول آن با طول پیام برابر است با پیام XOR می‌کند و متن رمزی $c = m \oplus k$ را تولید می‌کند. حال کلید تصادفی را با کلید "شبه‌تصادفی" جایگزین می‌کنیم؛ یعنی از یک مولد شبه‌تصادفی $G(\cdot)$ استفاده می‌کنیم که کلید k با طول کم (مثلاً ۸۰ بیت) را به دنباله شبه‌تصادفی $z = G(k)$ با طول زیاد (مثلاً چند مگابیت) می‌برد و سپس از آن برای رمز کردن پیام m به صورت $c = m \oplus z$ استفاده می‌کنیم.

در این جلسه مفاهیم شبه‌تصادفی و مولد شبه‌تصادفی را تعریف می‌کنیم. در ابتدا لازم است که مفهوم مهم تمایزناپذیری را بیان کنیم.

۱ تمایزناپذیری

تعریف ۱ *می‌گوییم متغیرهای تصادفی X_0 و X_1، (t, ε)-تمایزناپذیراند، اگر برای هر مهاجم (تمایزگر[1]) مانند $\mathcal{A}$ که در زمان t اجرا می‌شود داشته باشیم:*

$$\Pr\{b \leftarrow \{0, 1\}; x \leftarrow X_b : \mathcal{A}(x) = b\} \leq \frac{1}{2}(1 + \varepsilon) \qquad (1)$$

قضیه زیر تعریفی معادل برای مفهوم تمایزناپذیری بیان می‌کند:

قضیه ۱ *متغیرهای تصادفی X_0 و X_1، (t, ε)-تمایزناپذیرند، اگر و فقط اگر برای هر تمایزگر $\mathcal{D}$ که در زمان t اجرا می‌شود داشته باشیم:*

$$|\Pr\{x \leftarrow X_0 : \mathcal{D}(x) = 1\} - \Pr\{x \leftarrow X_1 : \mathcal{D}(x) = 1\}| \leq \varepsilon \qquad (2)$$

برهان. به طور شهودی تعریف ۱ می‌گوید که هیچ تمایزگری نمی‌تواند در زمان t با احتمالی خیلی بهتر از یک حدس تصادفی تعیین کند که x داده شده از کدام توزیع آمده است. این تعریف در واقع بیان می‌کند که هر تمایزگری که در زمان t اجرا می‌شود، با دیدن یک خروجی از توزیع X_0 و یک خروجی از توزیع X_1 تقریباً به طور یکسان عمل

[1] distinguisher

می‌کند. حال چون این تمایزگر یک بیت خروجی می‌دهد، "تقریباً یکسان عمل کردن" به معنای این است که ۱ را در هر مورد با احتمال تقریباً برابری خروجی می‌دهد و این همان تعریف ۲ است.
برای اثبات دقیق ابتدا نشان می‌دهیم که تعریف ۱، تعریف ۲ را نتیجه می‌دهد. فرض خلف کنید که چنین نباشد و تمایزگر $\mathcal{D}$ وجود داشته باشد که برای آن داشته باشیم:

$$|\Pr\{x \leftarrow X_۰ : \mathcal{D}(x) = ۱\} - \Pr\{x \leftarrow X_۱ : \mathcal{D}(x) = ۱\}| > \varepsilon$$

می‌توان فرض کرد (چرا؟):

$$\Pr\{x \leftarrow X_۱ : \mathcal{D}(x) = ۱\} - \Pr\{x \leftarrow X_۰ : \mathcal{D}(x) = ۱\} > \varepsilon$$

حال تمایزگر $\mathcal{A}$ را بدین صورت در نظر بگیرید: $\mathcal{A}$ با دریافت ورودی x آن را به $\mathcal{D}$ می‌دهد. سپس خروجی خود را بدین صورت بیرون می‌دهد:

$$\mathcal{A}(x) = \begin{cases} ۱ & \text{if } \mathcal{D}(x) = ۱ \\ ۰ & \text{if } \mathcal{D}(x) \neq ۱ \end{cases}$$

داریم:

$$\begin{aligned}
\Pr\{b \leftarrow \{۰,۱\}; x \leftarrow X_b : \mathcal{A}(x) = b\} &= \frac{۱}{۲}\Pr\{x \leftarrow X_۰ : \mathcal{A}(x) = ۰\} \\
&\quad + \frac{۱}{۲}\Pr\{x \leftarrow X_۱ : \mathcal{A}(x) = ۱\} \\
&= \frac{۱}{۲}\Pr\{x \leftarrow X_۰ : \mathcal{D}(x) \neq ۱\} \\
&\quad + \frac{۱}{۲}\Pr\{x \leftarrow X_۱ : \mathcal{D}(x) = ۱\} \\
&= \frac{۱}{۲}(۱ - \Pr\{x \leftarrow X_۰ : \mathcal{D}(x) = ۱\}) \\
&\quad + \frac{۱}{۲}\Pr\{x \leftarrow X_۱ : \mathcal{D}(x) = ۱\} \\
&= \frac{۱}{۲} + \frac{۱}{۲}(\Pr\{x \leftarrow X_۱ : \mathcal{D}(x) = ۱\} \\
&\quad - \Pr\{x \leftarrow X_۰ : \mathcal{D}(x) = ۱\}) \\
&\geq \tfrac{۱}{۲}(۱ + \varepsilon)
\end{aligned}$$

برای اثبات اینکه تعریف ۲، تعریف ۱ را نتیجه می‌دهد، فرض خلف کنید که چنین نباشد و تمایزگر $\mathcal{A}$ وجود داشته باشد که برای آن داشته باشیم:

$$\Pr\{b \leftarrow \{۰,۱\}; x \leftarrow X_b : \mathcal{A}(x) = b\} \geq \frac{۱}{۲}(۱ + \varepsilon)$$

تمایزگر $\mathcal{D}$ همان $\mathcal{A}$ در نظر بگیرید. داریم:

$$\begin{aligned}
&|\Pr\{x \leftarrow X_۰ : \mathcal{D}(x) = ۱\} - \Pr\{x \leftarrow X_۱ : \mathcal{D}(x) = ۱\}| \\
=\ &|\Pr\{x \leftarrow X_۰ : \mathcal{A}(x) = ۱\} - \Pr\{x \leftarrow X_۱ : \mathcal{A}(x) = ۱\}| \\
=\ &|۱ - \Pr\{x \leftarrow X_۰ : \mathcal{A}(x) = ۰\} - \Pr\{x \leftarrow X_۱ : \mathcal{A}(x) = ۱\}| \\
=\ &|۱ - ۲\Pr\{b \leftarrow \{۰,۱\}; x \leftarrow X_b : \mathcal{A}(x) = b\}| \\
\geq\ &\varepsilon
\end{aligned}$$

■

تعریف ۲ *می‌گوییم تمایزگر $\mathcal{D}$ توزیع‌های $X_۰$ و $X_۱$ را با احتمال (مزیت[۲]) μ تمایز می‌دهد اگر:*

$$|\Pr\{a \longleftarrow X_۰ : \mathcal{D}(a) = ۱\} - \Pr\{a \longleftarrow X_۱ : \mathcal{D}(a) = ۱\}| \geqslant \mu$$

مثال ۳ *فرض کنید متغیر تصادفی $X_۱$ دارای توزیع $\Pr\{X_۱ = ۱\} = ۳/۴$ و $\Pr\{X_۱ = ۰\} = ۱/۴$ روی مجموعه $\{۰, ۱\}$ باشد و متغیر تصادفی $X_۰$ دارای توزیع یکنواخت روی همان مجموعه باشد. مزیت تمایزگر $\mathcal{D}_۱$ که همواره خروجی ۱ را برمی گرداند، صفر است. همچنین تمایزگر تصادفی $\mathcal{D}_۲$ که مقادیر ۱ و ۰ را با احتمال $\frac{۱}{۲}$ بر می‌گرداند، نیز دارای مزیت صفر است. اما تمایزگر $\mathcal{D}_۳$ که به صورت زیر تصمیم می‌گیرد، دارای مزیت $\frac{۱}{۴}$ است.*

$$\mathcal{D}_۳(a) = \begin{cases} ۱ & a = ۱ \\ ۰ & a = ۰ \end{cases}$$

قضیه ۲ *بیشترین مقدار مزیتی که یک تمایزگر برای تمایز دو توزیع داده شده $X_۰$ و $X_۱$ می‌تواند داشته باشد برابر است با:*

$$\varepsilon = \frac{۱}{۲} \sum_x |\Pr[X_۱ = x] - \Pr[X_۰ = x]| .$$

که همان فاصله‌ی آماری[۳] بین دو توزیع می‌باشد. تمایزگر $\mathcal{D}$ است که به صورت زیر عمل می‌کند، دارای مزیت بیشینه است:

$$\mathcal{D}(a) = \begin{cases} ۱ & \Pr\{X_۱ = a\} \geq \Pr\{X_۰ = a\} \\ ۰ & \textit{oth.} \end{cases}$$

تعریف ۴ *تابع $G : \{۰, ۱\}^n \longrightarrow \{۰, ۱\}^l$ یک مولد شبه‌تصادفی(t, ε)-امن[۴] است اگر:*

- *به صورت کارایی قابل محاسبه باشد.*
- *طول خروجی بیشتر از ورودی باشد (ورودی را توسعه دهد).*
- *توزیع‌های U_l و $G(U_n)$، (t, ε)-تمایزناپذیر باشند.*

۲ رمز دنباله‌ای

تعریف ۵ *فرض کنید که $G : \{۰, ۱\}^n \to \{۰, ۱\}^l$ یک مولد شبه‌تصادفی باشد. در اینصورت رمز دنباله‌ای $\Pi_G = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ روی فضای پیام $\{۰, ۱\}^l$ با الگوریتم‌های زیر تعریف می‌شود:*

- Gen*: الگوریتم تولید کلید، کلید k را به تصادف از $\{۰, ۱\}^n$ انتخاب می‌کند.*

[۲] advantage
[۳] statistical distance
[۴] Pseudo Random Generator

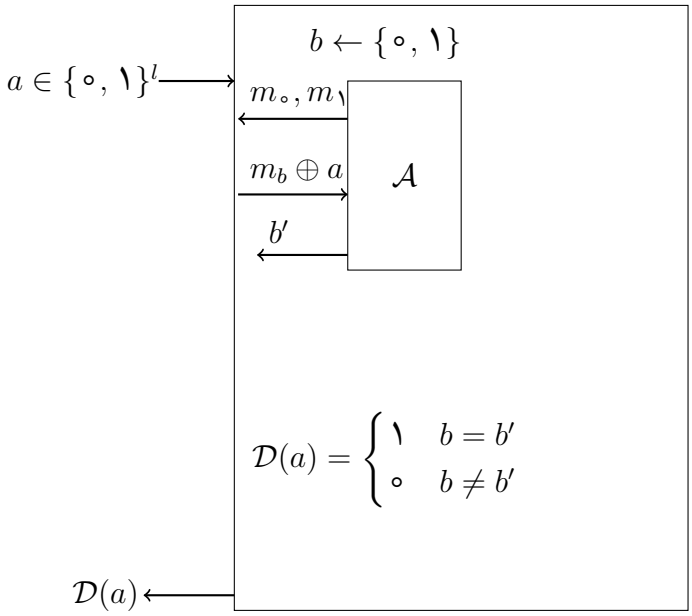
- Enc: *الگوریتم رمزنگاری، پیام* $m \in \{0,1\}^l$ *و کلید* $k \in \{0,1\}^n$ *را به متن رمز شده* $c = m \oplus G(k)$ *تبدیل می‌کند.*
- Dec: *الگوریتم رمزگشایی، متن رمز شده* $c \in \{0,1\}^l$ *و کلید* $k \in \{0,1\}^n$ *را به پیام* $m = c \oplus G(k)$ *تبدیل می‌کند.*

قضیه ۳ *اگر مولد شبه‌تصادفی* G *یک مولد شبه‌تصادفی* $(t, \varepsilon/2)$*-امن باشد رمز دنباله‌ای* Π_G *یک سیستم رمز متقارن* $(t - O(l), \varepsilon)$*-امن است.*

برهان. قرار می‌دهیم $t' = t - O(l)$ با استفاده از برهان خلف، درستی حکم را نشان می‌دهیم: پس باید نشان دهیم اگر رمز دنباله‌ای (t', ε)-امن نباشد آنگاه G، $(t' + O(l), \varepsilon)$-امن نیست. اگر رمز دنباله‌ای (t', ε)-امن نباشد پس مهاجمی مانند $\mathcal{A}$ وجود دارد که در زمان حداکثر t' کار می‌کند و با مزیت حداقل ε در آزمایش $\mathsf{PrivK}^{\mathrm{eav}}_{\mathcal{A},\Pi_G}$ موفق می‌شود؛ یعنی،

$$\Pr\{\mathsf{PrivK}^{\mathrm{eav}}_{\mathcal{A},\Pi_G} = 1\} \geq \frac{1}{2}(1 + \varepsilon) .$$

حال با استفاده از این حمله‌کننده به صورت زیر تمایزگری به نام $\mathcal{D}$ می‌سازیم.



تمایزگر $\mathcal{D}$ نقش چالشگر را برای مهاجم $\mathcal{A}$ شبیه‌سازی می‌کند. به این صورت که تمایزگر $\mathcal{D}$ پس از دریافت رشته l بیتی a به عنوان ورودی، حمله‌کننده $\mathcal{A}$ را اجرا می‌کند و دو متن چالش m_0 و m_1 را از آن می‌گیرد. تمایزگر $\mathcal{D}$ سپس بیت b را به تصادف انتخاب می‌کند و حاصل $m_b \oplus a$ را به عنوان متن رمزی به حمله‌کننده می‌دهد. در مرحله بعد حمله‌کننده باید b' را به عنوان خروجی به تمایزگر بدهد، اگر $b = b'$ تمایزگر عدد ۱ را بر می‌گرداند درغیر این صورت ۰. از طرفی ورودی a یا خروجی مولد شبه‌تصادفی است یا کاملا تصادفی است. اگر ورودی a کاملا تصادفی باشد، داریم:

$$\Pr\{a \leftarrow U_l : \mathcal{D}(a) = 1\} = \frac{1}{2}$$

اگر ورودی a، خروجی مولد شبه‌تصادفی باشد، در این صورت خروجی تمایزگر $\mathcal{D}$ در واقع همان خروجی آزمایش $\mathsf{PrivK}^{\mathrm{eav}}_{\mathcal{A},\Pi_G}$. بنابراین خواهیم داشت:

$$\Pr\{k \leftarrow \{۰,۱\}^n; a \leftarrow G(k) : \mathcal{D}(a) = ۱\} = \Pr\{\mathsf{PrivK}^{\mathrm{eav}}_{\mathcal{A},\Pi_G} = ۱\} \geq \frac{۱}{۲}(۱ + \varepsilon)$$

پس داریم :

$$|\Pr\{a \leftarrow U_l : \mathcal{D}(a) = ۱\} - \Pr\{k \leftarrow \{۰,۱\}^n; a \leftarrow G(k) : \mathcal{D}(a) = ۱\}| \geq \varepsilon/۲$$

چون $\mathcal{A}$ حداکثر در زمان t' اجرا می‌شود پس $\mathcal{D}$ حداکثر در زمان $t' + O(l)$ انجام می‌شود، زیرا تنها کار اضافه‌تری که انجام می‌دهد عمل xor رشته l-بیتی است. بنابراین فرض خلف باطل است و در نتیجه سیستم رمز دنباله‌ای $(t - O(l), \varepsilon)$-امن است. ■

۳ چگونه مولد شبه‌تصادفی بسازیم؟

ایده‌های ساده‌ای مثل تکرار کردن ورودی به منظور گسترش آن، مولد شبه‌تصادفی نیستند.

مثال ۶ *مولد $G : \{۰,۱\}^n \to \{۰,۱\}^{۲n}$ را در نظر بگیرید که رشته n-بیتی k را به رشته $۲n$-بیتی $k\|k$ می‌برد، این مولد شبه‌تصادفی نیست. تمایزگر $\mathcal{D}$ را که به صورت زیر عمل می‌کند در نظر بگیرید که $z_۱, z_۲ \in \{۰,۱\}^n$:*

$$\mathcal{D}(z_۱\|z_۲) = \begin{cases} ۱ & z_۱ = z_۲ \\ ۰ & z_۱ \neq z_۲ \end{cases}$$

داریم:

$$\Pr\{k \longleftarrow U_n; z = G(k) : \mathcal{D}(z) = ۱\} = ۱$$
$$\Pr\{z \longleftarrow U_{۲n} : \mathcal{D}(z) = ۱\} = ۲^{-n}$$

پس داریم:
$|\Pr\{k \longleftarrow U_n; z \leftarrow G(k) : \mathcal{D}(z) = ۱\} - \Pr\{z \longleftarrow U_{۲n} : \mathcal{D}(z) = ۱\}| = ۱ - ۲^{-n}$
تمایزگر می‌تواند مولد را با مزیت قابل توجه تمایز دهد و در نتیجه این مولد، شبه‌تصادفی نیست.

برای ساخت مولدهای شبه‌تصادفی باید از ایده‌هایی استفاده کنیم که بیت‌های ورودی را به طور "پیچیده"ای با هم مخلوط می‌کنند. در جلسه بعد، LFSRها را که دنباله خروجی‌شان دارای خواص خوبی است، بررسی خواهیم کرد. سپس به مطالعه روش‌های طراحی مولدهای شبه تصادفی با استفاده از LFSRها خواهیم پرداخت.