به نام خدا

دانشکده‌ی علوم ریاضی

مقدمه‌ای بر رمزنگاری ۱۵ اسفند ۱۳۹۱

جلسه‌ی ۹: توابع بولی و تحلیل مولد فیلتری

مدرس: دکتر شهرام خزائی *نگارنده: محسن صحرائی*

۱ توابع بولی و تعاریف

تعریف ۱ *هر تابع* $f : \{۰, ۱\}^n \longrightarrow \{۰, ۱\}$ *یک تابع بولی*[۱] *(n-متغیره) نامیده می‌شود.*

دو روش برای نمایش یک تابع بولی وجود دارد.

- نمایش مقداری: یک تابع بولی f بر روی $\{۰, ۱\}^n$ را می‌توان با استفاده از جدول صحت[۲] به صورت یکتا مشخص کرد. جدول صحت برداری به صورت $[f(\langle ۰ \rangle), \cdots, f(\langle ۲^{n-۱} \rangle)]$، شامل همه‌ی مقادیر تابع f است که $\langle i \rangle$ بیان‌گر رشته باینری به طول n است که نمایش مبنای دو عدد i است.
- نمایش جبری: روش دیگری که می‌توان با استفاده از آن یک تابع بولی را به صورت یکتا مشخص کرد، استفاده از فرم نرمال جبری[۳] است. فرم نرمال جبری یک چندجمله‌ای به صورت زیر است:

$$f(x) = \sum a_I \prod_{i \in I} x_i \,, \qquad I \subseteq \{۱, ۲, ۳, \cdots, n\},\ a_I \in \{۰, ۱\}$$

مثال ۲ *عبارت زیر نمایش یک تابع بولی دومتغیره با استفاده از فرم نرمال جبری است:*

$$f(x) = a_۰ + a_۱x_۱ + a_۲x_۲ + a_۳x_۱x_۲, \qquad a_i \in \{۰, ۱\}$$

جدول صحت این تابع بولی به صورت زیر است:

$x_۱$	$x_۲$	$f(x)$
۰	۰	$a_۰$
۰	۱	$a_۰ + a_۲$
۱	۰	$a_۰ + a_۱$
۱	۱	$a_۰ + a_۱ + a_۲ + a_۳$

[۱] Boolean Function
[۲] Truth Table
[۳] Algebraic Normal Form

نکته ۱ *بین دو روش جدول صحت و فرم نرمال جبری، می‌توان با مرتبه‌ی پیچیدگی $n\mathbf{2}^n$ تبدیل انجام داد.*

برای ادامه‌ی بحث به تعاریفی نیاز است که در ادامه به آنها خواهیم پرداخت.

تعریف ۳ *درجه‌ی جبری[۴] تابع بولی f که با $\deg(f)$ نمایش داده می‌شود، بیشینه‌ی تعداد متغیرهای موجود در عبارت‌های سازنده‌ی فرم نرمال جبری است؛ یعنی:*

$$\deg(f) = \max\{|I|, a_I \neq \circ\}$$

تعریف ۴ *تابع مستوی[۵]، تابعی است که درجه‌ی جبری آن کمتر یا مساوی یک باشد.*

به عبارت دیگر، تابعی به فرم زیر را تابع مستوی می‌گویند:

$$f(x) = a_\circ + a_\text{۱}x_\text{۱} + \cdots + a_n x_n$$

در عبارت فوق، اگر $a_\circ = \circ$ برقرار باشد، آنگاه تابع را خطی[۶] می‌گویند.

تعریف ۵ *وزن همینگ[۷] تابع f، تعداد مقادیر غیرصفر تابع f به ازای تمام ورودی‌های ممکن است؛ یعنی:*

$$w_H = |\{x \in \{\circ, \text{۱}\}^n \mid f(x) \neq \circ\}|$$

در حقیقت وزن همینگ تابع f، برابر تعداد ۱ها در جدول صحت آن است.

تعریف ۶ *تابع متوازن[۸]، تابعی است که تعداد مقادیر صفر و یک در جدول صحت آن برابر باشد.*

نکته ۲ *با توجه به تعریف فوق در توابع متوازن رابطه‌ی زیر برقرار است:*

$$w_H(f) = \mathbf{2}^{n-\text{۱}}$$

همچنین داریم:

$$\Pr_{x \leftarrow \{\circ,\text{۱}\}^n}\{f(x) = \circ\} = \frac{\text{۱}}{\mathbf{2}}$$

تعریف ۷ *فاصله‌ی همینگ[۹] دو تابع f و g که با $d_H(f,g)$ نمایش داده می‌شود، برابر است با تعداد مقادیر متفاوت f و g به ازای همه‌ی ورودی‌های ممکن.*

فاصله‌ی دو تابع را می‌توان از رابطه‌ی زیر به دست آورد:

$$d_H(f,g) = w_H(f+g)$$

[۴] Algebric Degree
[۵] Affine Function
[۶] Linear
[۷] Hamming Weight
[۸] Balanced Function
[۹] Hamming Distance

نکته ۳ *با توجه به تعریف فوق می‌توان گفت:*

$$\Pr_{x \leftarrow \{۰,۱\}^n} \{f(x) \neq g(x)\} = \frac{d_H(f,g)}{۲^n}$$

اگر تابعی را با تابع دیگر تقریب بزنیم، با استفاده از فاصله‌ی بین دو تابع می‌توان تعیین کرد که تقریب تا چه حد خوب است. امّا در بسیاری از موارد تقریب‌های خطی مطلوب ماست. به همین خاطر تعریف زیر را مطرح می‌کنیم.

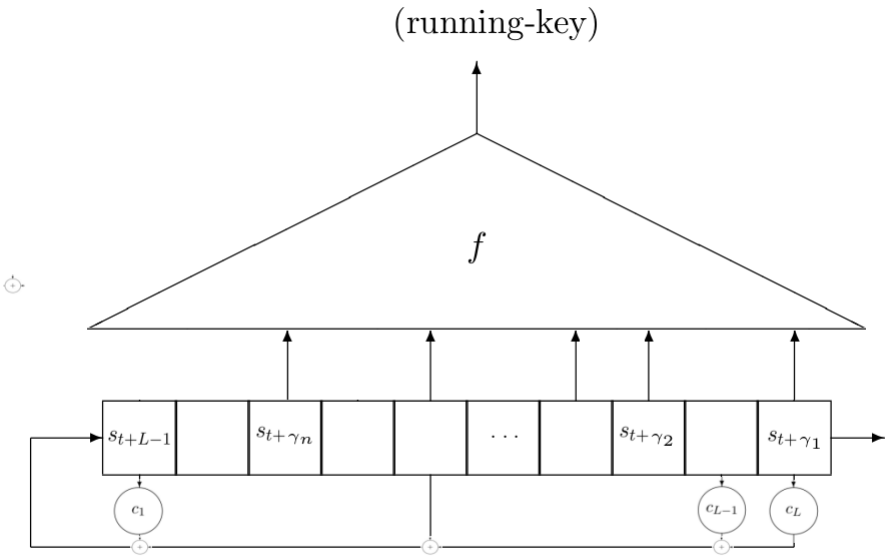
تعریف ۸ *مقدار غیرخطی*[۱۰] *تابع f را با N_f نمایش می‌دهند و برابر است با کمینه‌ی فاصله‌ی f از مجموعه‌ی توابع مستوی. به عبارت دیگر مقدار غیرخطی f برابر است با فاصله‌ی آن تا نزدیک‌ترین تابع مستوی.*

$$N_f = \min_{affine\ g's} \{d_H(f,g)\}$$

۲ تحلیل مولد فیلتری

فرض کنید یک مولد فیلتری با چندجمله‌ای ارتباط $X^L + \sum_{i=۰}^{L-۱} c_{L-i}x^i$، تابع فیلتر $f(x_۱, x_۲, \cdots x_n)$ و موقعیت‌های $\gamma_۱, \gamma_۲, \cdots, \gamma_n$ که $۰ \leq \gamma_۱ < \gamma_۲ < \cdots < \gamma_n \leq L-۱$ داریم. دنباله‌ی خروجی این مولد که با $\{z_t\}_{t\geq ۰}$ نشان داده می‌شود از روی دنباله‌ی خروجی ثبات خطی که با $\{s_t\}_{t\geq ۰}$ نشان داده می‌شود از رابطه‌ی زیر به‌دست می‌آید:

$$z_t = f\left(s_{t+\gamma_۱}, s_{t+\gamma_۲}, \cdots, s_{t+\gamma_n}\right)$$



[۱۰] Nonlinearity

به یاد آورید که رابطه‌ی خطی زیر برای دنباله‌ی خروجی ثبات خطی برای $t \geq L$ برقرار است که $c_۰ = ۱$ در نظر گرفته می‌شود:

$$\sum_{i=۰}^{L} c_i s_{t-i} = ۰$$

در ادامه حمله‌ی تمایز و حمله‌ی حالت اولیه را برای f خطی و غیرخطی بررسی می‌کنیم.

۱.۲ حمله‌ی تمایز

در حمله‌ی تمایز، امنیت سیستم با توانایی تشخیص خروجی آن از یک دنباله‌ی تصادفی محک زده می‌شود. این حمله را برای رمز دنباله‌ای با فیلتر خطی و غیرخطی بررسی می‌کنیم.

۱.۱.۲ مولد فیلتری با فیلتر خطی

با توجه به رابطه‌ای که در قسمت قبل گفته شد، می‌توان لم زیر را مطرح کرد.

لم ۱ *اگر f خطی باشد داریم:*

$$\sum_{i=۰}^{L} c_i z_{t-i} = ۰$$

برهان. فرض کنید تابع فیلتر خطی به صورت $f(x_۱, \cdots, x_n) = \sum_{j=۱}^{n} a_j x_j$ باشد داریم:

$$z_t = \sum_{j=۱}^{n} a_j s_{t+\gamma_j}$$

$$\begin{aligned}
\sum_{i=۰}^{L} c_i z_{t-i} &= \sum_{i=۰}^{L} c_i \sum_{j=۱}^{n} a_j s_{t-i-\gamma_j} \\
&= \sum_{j=۱}^{n} a_j \sum_{i=۰}^{L} c_i s_{t-i-\gamma_j} \\
&= \sum_{j=۱}^{n} a_j (۰) = ۰
\end{aligned}$$

■

۲.۱.۲ مولد فیلتری با فیلتر غیرخطی

اگر تابع f غیرخطی باشد لم ۱ دیگر برقرار نیست. در این حالت با داشتن یک تقریب خطی $a_1x_1 + \cdots + a_nx_n$ برای تابع غیرخطی f می‌توان نوشت:

$$\Pr\{f(x) = a_1x_1 + \cdots + a_nx_n\} = \frac{1}{2}(1 + \varepsilon), \qquad \varepsilon \neq 0$$

یعنی یک تقریب خطی برای تابع خود داریم. بنابراین می‌توان نوشت:

$$f(x) = a_1x_1 + \cdots + a_nx_n + e$$

که e یک متغیر تصادفی باینری با توزیع زیر است که آنرا نویز می‌نامیم:

$$\Pr\{e = 0\} = \frac{1}{2}(1 + \varepsilon)$$

بنابراین:

$$z_t = \left(\sum_{j=1}^{n} a_j s_{t+\gamma_j}\right) + e_t$$

که $\{e_t\}$ دنباله نویز می‌باشد.

لم ۲ *فرض کنید که $a_1x_1 + \cdots + a_nx_n$ یک تقریب خطی دلخواه برای تابع فیلتر f باشد. فرض کنید که چندجمله‌ای مشخصه‌ی LFSR به صورت $x^{i_1} + x^{i_2} + \cdots x^{i_w}$ باشد، در این صورت*

$$\sum_{i=0}^{L} c_i z_{t-i} = e_{t-i_1} + \cdots + e_{t-i_w}$$

برهان. داریم:

$$\begin{aligned}
\sum_{i=0}^{L} c_i z_{t-i} &= \sum_{i=0}^{L} c_i f(s_{t-i+\gamma_1}, \cdots, s_{t-i+\gamma_n}) \\
&= \sum_{i=0}^{L} c_i \left(\sum_{j=1}^{n} a_j s_{t-i+\gamma_j} + e_{t-i}\right) \\
&= \sum_{i=0}^{L} c_i \sum_{j=1}^{n} a_j s_{t-i+\gamma_j} + \sum_{i=0}^{L} c_i e_{t-i} \\
&= 0 + \sum_{i=0}^{L} c_i e_{t-i} \\
&= e_{t-i_1} + \cdots + e_{t-i_w}
\end{aligned}$$

■

لم ۳ (لم pilling-up) *فرض کنید n متغیر تصادفی باینری مستقل $X_۱, \cdots, X_n$ با توزیع احتمال زیر داریم:*

$$\Pr(X_i = ۱) = \frac{۱}{۲}(۱ - \varepsilon_i)$$

$$\Pr(X_i = ۰) = \frac{۱}{۲}(۱ + \varepsilon_i)$$

در این‌صورت متغیر تصادفی

$$X = X_۱ + \cdots + X_n$$

دارای توضیح احتمال زیر است:

$$\Pr(X = ۱) = \frac{۱}{۲}(۱ - \varepsilon_۱ \cdots \varepsilon_n)$$

$$\Pr(X = ۰) = \frac{۱}{۲}(۱ + \varepsilon_۱ \cdots \varepsilon_n)$$

برهان. این لم را برای دو متغیر تصادفی اثبات می‌کنیم. برای تعداد بیشتر با استقراء ثابت می‌شود. بنابراین فرض می‌کنیم حالت زیر را داریم:

	۰	۱
$X_۱$	$\frac{۱}{۲}(۱ + \varepsilon_۱)$	$\frac{۱}{۲}(۱ - \varepsilon_۱)$
$X_۲$	$\frac{۱}{۲}(۱ + \varepsilon_۱)$	$\frac{۱}{۲}(۱ - \varepsilon_۲)$

$$\begin{aligned}
\Pr\{X = ۱\} &= \Pr\{X_۱ + X_۲ = ۱\} \\
&= \Pr\{X_۱ \neq X_۲\} \\
&= \frac{۱}{۲}(۱ - \varepsilon_۱)\frac{۱}{۲}(۱ + \varepsilon_۲) + \frac{۱}{۲}(۱ + \varepsilon_۱)\frac{۱}{۲}(۱ - \varepsilon_۲) \\
&= \frac{۱}{۲}(۱ - \varepsilon_۱\varepsilon_۲)
\end{aligned}$$

$$\begin{aligned}
\Pr\{X = ۰\} &= \Pr\{X_۱ + X_۲\} \\
&= \Pr\{X_۱ = X_۲\} \\
&= \frac{۱}{۲}(۱ + \varepsilon_۱)\frac{۱}{۲}(۱ + \varepsilon_۲) + \frac{۱}{۲}(۱ - \varepsilon_۱)\frac{۱}{۲}(۱ - \varepsilon_۲) \\
&= \frac{۱}{۲}(۱ + \varepsilon_۱\varepsilon_۲)
\end{aligned}$$

■

به طور خلاصه به فرض مستقل بودن e_tها (که در عمل تا حد زیادی صادق است)، داریم:

$$\Pr\left\{\sum_{i=0}^{L} c_i z_{t-i} = 0\right\} = \frac{1}{2}\left(1 + \varepsilon^w\right)$$

که w تعداد ضرایب ناصفر چندجمله‌ای مشخصه (ارتباط) ثبات خطی است و ε بایاس بهترین تقریب خطی تابع فیلتر است.
اگر Y متغیر تصادفی کاملاً تصادفی باشد، خواهیم داشت:

$$\Pr(Y = 1) = \Pr(Y = 0) = \frac{1}{2}$$

بنابراین با استفاده از یک نمونه بین X و Y با احتمال (مزیت) ε^w می‌توانیم تمیز دهیم.
اگر تعداد نمونه‌ها $N = \frac{1}{\varepsilon^{2w}}$ باشد، مزیت تمایزگر به یک نزدیک می‌شود.

مثال ۹

$$L = 100, \quad w = 5, \quad , \varepsilon = 2^{-3}$$

$$N \simeq \frac{1}{\varepsilon^{2w}} = 2^{30}$$

با 2^{30} نمونه می‌توان این رمز دنباله‌ای را از رشته‌ی کاملاً تصادفی تمیز داد.

در حالت کلی اگر دو عبارت زیر برقرار باشند:

$$\Pr\{X = 0\} = p\,(1 + \varepsilon) \quad (1)$$
$$\Pr\{Y = 0\} = p \quad (2)$$

آنگاه تعداد نمونه‌های لازم برای تشخیص دو متغیر تصادفی به صورت زیر به دست خواهد آمد:

$$N = \frac{1}{p\varepsilon^2}$$

۲.۲ حمله کشف حالت اولیه

در این نوع حمله رابطه‌ی مولد فیلتری و نیز خروجی را در اختیار داریم و می‌خواهیم حالت اولیه‌ی LFSR را بیابیم.
برای این کار از دو روش می‌توان استفاده کرد:

۱. روش جبری

۲. روش همبستگی

در اینجا تمرکز ما بر روی روش جبری است و به روش همبستگی نخواهیم پرداخت. این قسمت را نیز به صورت اجمالی در دو بخش توابع خطی و غیرخطی بررسی می‌کنیم.

۱.۲.۲ مولد فیلتری با فیلتر خطی

اگر فیلتر خطی باشد، می‌توان از روی L بیت دنباله خروجی مولد، L معادله بر حسب L مجهول حالت اولیه نوشت و (مشروط بر اینکه معادلات مستقل خطی باشند) با حل آن حالت اولیه را یافت . در این حالت دستگاه معادلات با استفاده از روش حذفی گاوس در $O(L^۳)$ حل می‌شود.

۲.۲.۲ مولد فیلتری با فیلتر غیرخطی

اگر فیلتر غیرخطی باشد می‌توان از تکنیک خطی‌سازی[۱۱] استفاده کرد. بدین معنی که به جای هر جمله به صورت $\Pi_{i\in I}s_i$ که $I \subseteq \{۰, ۱, \ldots, L-۱\}$ یک متغیر جدید به نام y_I معرفی می‌کنیم. بنابراین هر بیت دنباله خروجی مولد را می‌توان به صورت ترکیبی خطی از متغیرهای خطی‌سازی‌شده y_I نوشت. توجه کنید که اگر درجه‌ی تابع فیلتر غیرخطی d باشد فقط متغیرهای خطی‌سازی شده y_I که $|I| \leq d$ در معادلات خطی‌سازی‌شده ظاهر خواهند شد. بنابراین تعداد متغیرهای خطی‌سازی شده برابر $\sum_{i=۰}^{d}\binom{L}{i}$ خواهد بود که با استفاده از روش حذفی گاؤس در زمان $\left(\sum_{i=۰}^{d}\binom{L}{i}\right)^۳$ قابل حل است.

[۱۱] linearization