به نام خدا

دانشکده‌ی علوم ریاضی

مقدمه‌ای بر رمزنگاری ۱۷ بهمن ۱۳۹۱

جلسه‌ی ۲: سیستم رمز متقارن و مدل‌های حمله

مدرس: دکتر شهرام خزائی *نگارنده: امیربهشاد شهراسبی*

۱ اصل کرشهف

اصل کرشهف[۱] پذیرفتن فرض‌های زیر را در سیستم‌های رمزنگاری مطرح می‌کند:

۱. الگوریتم رمزنگاری آشکار است.

۲. امنیّت تنها به مخفی بودن کلید وابسته است.

۲ سیستم رمز متقارن

تعریف ۱ *یک سیستم رمز متقارن[۲]، یک سه‌تایی* $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *از الگوریتم‌های کاراست که:*

- Gen *الگوریتم تولید کلید است که یک کلید* k *از فضای کلید* $\mathcal{K}$ *تولید و فضای پیام* $\mathcal{M}$ *را مشخص می‌کند:*

$$k \leftarrow \mathsf{Gen}()$$

- Enc *الگوریتم رمزنگاری است که متن اصلی* $m \in \mathcal{M}$ *و کلید* $k \in \mathcal{K}$ *را به متن رمزی* c *تبدیل می‌کند.*

$$c \leftarrow \mathsf{Enc}_k(m)$$

- Dec *الگوریتم رمزگشایی است که هر مقدار* c *و کلید* $k \in \mathcal{K}$ *را به متن اصلی* m *یا* $m = \perp \notin \mathcal{M}$ *تبدیل می‌کند.*

$$m \leftarrow \mathsf{Dec}_k(c)$$

- **(شرط صحت)** *به ازای هر* $m \in \mathcal{M}$، $k \in \mathcal{K}$ *و* $c \in \mathcal{C}$ *که* $c = \mathsf{Enc}_k(m)$ *داریم* $\mathsf{Dec}_k(c) = m$؛ *به عبارت دیگر:*

[۱] Kerckhoffs's principle
[۲] symmetric cipher

$$\forall m \in \mathcal{M} \quad \Pr\{k \leftarrow \mathsf{Gen}() : \mathsf{Dec}_k(\mathsf{Enc}_k(m)) = m\} = ۱$$

که احتمال روی سکه‌های تصادفی الگوریتم‌ها محاسبه می‌شوند.

در اینجا ذکر چند نکته ضروری است:

- برای اینکه یک سیستم رمزنگاری در عمل قابل استفاده باشد،باید الگوریتم‌های Gen، Enc و Dec "کارا" باشند.
- تعریف فوق هیچ مفهومی از امنیت در خود ندارد. با این وجود به وضوح اگر الگوریتم تولید کلید تصادفی[۳] نباشد، سیستم دارای امنیت (به هر صورت مقبولی که تعریف شود) نخواهد بود.
- الگوریتم رمزنگاری معمولاً تصادفی است. اما سیستم‌هایی که در ابتدا معرفی و بررسی می‌شوند همگی دارای الگوریتم رمزنگاری قطعی[۴] هستند.
- الگوریتم رمزگشایی قطعی است. همچنین در صورتی که الگوریتم رمزگشایی نماد $\perp$ به عنوان خروجی برای رمز گشایی مقدار c تحت کلید k تولید کند، بدین معناست که متن رمزی معتبر نیست؛ به عبارت دیگر، مقدار c نمی‌تواند رمزشده‌ی هیچ متن اصلی تحت کلید k باشد.

تعریف ۲ *سیستم رمز سزار[۵] را به شکل زیر تعریف می‌کنیم:*

$$\mathcal{M} = \{a, b, \cdots, z\}^*$$

- تعریف Gen*: یک عدد تصادفی از مجموعه‌ی* $\mathcal{K} = \{۰, ۱, \cdots, ۲۵\}$ *تولید می‌کند.*
- تعریف Enc*:*

$$\mathsf{Enc}_k(m_۱ m_۲ \cdots m_t) = c_۱ c_۲ \cdots c_t$$
$$c_i = m_i + k \quad i = ۱, ۲, \cdots, t$$

- تعریف Dec*:*

$$\mathsf{Dec}_k(c_۱ c_۲ \cdots c_t) = m_۱ m_۲ \cdots m_t$$
$$m_i = c_i - k \quad i = ۱, ۲, \cdots, t$$

در واقع سیستم رمز سزار یک کلید تصادفی در $\mathcal{K}$ انتخاب می‌کند و حروف متن را به آن اندازه شیفت می‌دهد.
توجه کنید الگوریتم Enc در این سیستم قطعی[۶] است. حال آنکه در حالت کلی این الگوریتم می‌تواند احتمالاتی باشد. (بدین معنا که برای یک ورودی مشخّص، خروجی یک متغیّر تصادفی[۷] است و مقادیر مختلفی را با یک توزیع احتمالاتی معیّن اختیار می‌کند.)

[۳] probabilistic
[۴] determinisitc
[۵] Caesar cipher
[۶] deterministic
[۷] random variable

۳ مدل‌های حمله

مدل‌ها یا سناریوهای حمله، قابلیت‌های مهاجم را برای حمله به یک سیستم رمز در نظر می‌گیرد.
مدل‌ها (یا سناریوهای) حمله قدرت و توان مهاجم را در دستیابی به اطلاعات بیشتر (مانند استفاده از دستگاه رمزکننده و دستگاه رمزگشا) و میزان آن مدل می‌کند.

- **حمله‌ی متن رمزشده**[۸]: در این مدل، حمله کننده تنها یک متن رمز شده را دارد و می‌خواهد متن اصلی را رمزگشایی کند. این رمزگشایی تنها زمانی میسّر است که متن اصلی دارای افزونگی[۹] باشد.

- **حمله‌ی متن اصلی معلوم**[۱۰]: در این مدل، حمله کننده یک یا چند زوج (متن اصلی، متن رمز شده) که تحت یک کلید حاصل شده‌اند را دارد و می‌خواهد متن اصلی متناظر با یک متن رمزشده‌ی دیگر را بداند.

- **حمله‌ی متن اصلی انتخابی**[۱۱]: در این مدل، حمله کننده قادر به بدست آوردن متن رمزشده‌ی متناظر با هر متن اصلی‌ای که بخواهد هست و می‌خواهد متن اصلی متناظر با یک متن رمزشده را بداند.

- **حمله‌ی متن رمزی انتخابی**[۱۲]: در این مدل، حمله کننده می‌خواهد متن اصلی متناظر با یک متن رمزشده را بداند و مضاف بر اینکه قادر است یک متن رمز شده متناظر با هر متن اصلی دلخواه را بدست آورد، قادر به بدست آوردن متن اصلی متناظر با هر متن رمزشده‌ای که بخواهد نیز هست (جز متنی که می‌خواهد رمزگشایی کند).

۴ هدف حمله‌کننده

در مدل‌های حمله‌ی بالا، هدف حمله کننده را بدست آوردن متن اصلی متناظر با یک متن رمز شده در نظر گرفتیم.
در حالت کلی حمله می‌تواند با اهداف مختلفی صورت پذیرد:

- بدست آوردن کلید
- محاسبه‌ی متن اصلی معادل یک متن رمز شده
- بدست آوردن مقداری اطّلاعات دربازه‌ی متن اصلی
- اعمال حمله‌ی تمایز[۱۳] (تشخیص اینکه کدامیک از دو متن اصلی دلخواه رمز شده‌است)

---

[۸] ciphertext-only attack
[۹] redundancy
[۱۰] known-plaintext attack
[۱۱] chosen-plaintext attack
[۱۲] chosen-ciphertext attack
[۱۳] distinguishing

۵ چند سیستم ساده و حمله متن رمز شده به آنها

۱.۵ حمله خودکار به رمز سزار

ایده‌ی حمله به رمز سزار استفاده از توزیع غیر یکنواخت حروف زبان انگلیسی است که اندازه‌گیری‌های آماری نشان می‌دهد به شکل زیر است:

جدول ۱: توزیع حروف زبان انگلیسی بدون احتساب فاصله

a	8.17%	n	6.75%
b	1.49%	o	7.51%
c	2.78%	p	1.92%
d	4.25%	q	0.09%
e	12.70%	r	5.99%
f	2.23%	s	6.33%
g	2.01%	t	9.06%
h	6.09%	u	2.76%
i	6.97%	v	0.98%
j	0.15%	w	2.36%
k	0.77%	x	0.15%
l	4.02%	y	1.97%
m	2.41%	z	0.07%

اگر در جدول فوق فراوانی سمبل iام را f_i بنامیم، خواهیم داشت:

$$\sum_{i=۰}^{۲۶} f_i^۲ = ۰/۰۶۵$$

با علم به این موضوع، حمله‌ی خودکار به رمز سزار را می‌توان به شکل زیر انجام داد:

۱. محاسبه‌ی فراوانی حرف‌های متن رمز شده

۲. محاسبه‌ی ضرایب انطباق[۱۴]: ضرایب انطباق را به شکل زیر تعریف می‌کنیم:

$$I_j = \sum_{i=۰}^{۲۶} f_i \times p_{i+j}$$

اگر j برابر با k باشد و متن به اندازه‌ی کافی بزرگ باشد، انتظار داریم $I_j \approx ۰/۰۶۵$.

۳. حدس ما برای k اندیس jای است که عبارت $|I_j - ۰/۰۶۵|$ را کمینه کند.

$$k = \arg\min_j \{|I_j - ۰/۰۶۵|\}$$

[۱۴] coincidence index

توجه کنید این روش، لزوما بهترین روش نیست، اما روشی برای حمله به سیستم رمز سزار است. به عنوان مثال استفاده از توزیع دوحرفی‌ها[۱۵] یا سه‌حرفی‌ها[۱۶] به جای تک‌حرفی‌ها[۱۷] می‌تواند منجرّ به حمله‌ی بهتری شود.

۲.۵ سیستم رمز جایگزینی

تعریف ۳ *سیستم رمز جایگزینی[۱۸] با الگوریتم‌های زیر روی فضای پیام* $\mathcal{M} = \{a, b, \cdots, z\}^*$ *تعریف می‌شود:*

- تعریف Gen*: یک جایگشت تصادفی از* $(۰, ۱, \cdots, ۲۵)$ *مثل* k *را تولید می‌کند.*
- تعریف Enc*:*

$$\mathsf{Enc}_k(m_۱ m_۲ \cdots m_t) = c_۱ c_۲ \cdots c_t$$
$$c_i = k(m_i) \quad i = ۱, ۲, \cdots, t$$

- تعریف Dec*:*

$$\mathsf{Dec}_k(c_۱ c_۲ \cdots c_t) = m_۱ m_۲ \cdots m_t$$
$$m_i = k^{-۱}(c_i) \quad i = ۱, ۲, \cdots, t$$

جهت حمله به این سیستم می‌توان فراوانی حروف متن رمزی را بدست آورد و حروف را بر حسب فراوانی مرتب نمود و حرف iام این لیست را رمز شده‌ی iامین حرف پر تکرار الفبا در نظر گرفت. اگر متن به اندازه‌ی کافی بزرگ باشد این روش، روش مناسبی جهت حمله است.

۳.۵ سیستم رمز ویژنر

تعریف ۴ *سیستم رمز ویژنر[۱۹] با الگوریتم‌های زیر روی فضای پیام* $\mathcal{M} = \{a, b, \cdots, z, -\}^*$ *تعریف می‌شود:*

- تعریف Gen*: در این سیستم* $\mathsf{Gen}()$ *یک عضو مانند* $k = (k_۰, k_۱, \cdots, k_{d-۱})$ *به‌تصادف از فضای کلید* $\mathcal{K} = \{۰, ۱, \cdots, ۲۶\}^d$ *تولید می‌کند.*
- تعریف Enc*:*

$$\mathsf{Enc}_k(m_۰ m_۱ \cdots m_{\ell-۱}) = c_۰ c_۱ \cdots c_{\ell-۱}$$
$$c_i = m_i + k_{i \bmod d}$$

- تعریف Dec*:*

[۱۵] digrams
[۱۶] trigrams
[۱۷] monograms
[۱۸] substitution cipher
[۱۹] Vigenère cipher

$$\mathsf{Dec}_k(c_0 c_1 \cdots c_{\ell-1}) = m_0 m_1 \cdots m_{\ell-1}$$

$$m_i = c_i - k_{i \bmod d}$$

در واقع در این سیستم، متن اصلی به پنجره‌های dتایی تقسیم شده و iامین حرف هر پنجره ($0 \le i \le d-1$)، k_i واحد جابجا می‌شود. دقت کنید برای حمله به این سیستم، اگر مقدار حدس زده شده برای d را $\hat{d}$ بنامیم و $\hat{d} = d$، آنگاه برای متن به اندازه‌ی کافی طولانی، توزیع حروف برای حرف‌های iام پنجره‌ها ($0 \le i \le \hat{d}-1$) برابر شیفت یافته‌ی توزیع حروف زبان انگلیسی می‌شود. امّا اگر $\hat{d} \neq d$ توزیع‌های فوق به سمت توزیع یکنواخت می‌رود. پس برای حمله به این سیستم می‌توان معیارهای Σf_i^2 را برای عناصر هم‌مکان در پنجره‌ها تعریف نمود. سپس برای مقادیر مختلف $\hat{d}$، معیارها را بررسی کرد. برای $\hat{d} = d$ مقدار تمام معیارها نزدیک ۰.۰۶۵ خواهد شد. بدین ترتیب d بدست می‌آید. پس از آن رمزگشایی برای جایگاه‌های iام ($0 \le i \le d-1$) از پنجره‌ها مانند سیستم سزار انجام می‌شود. از مزایای این سیستم این است که توزیع تک‌حرفی‌ها، دوحرفی‌ها و سه‌حرفی‌ها برای dهای بزرگ، تقریبا یکنواخت خواهد بود.