امید ریاضی نرخ پوشش برای ماتریس‌های هلمن

ناصرحسین غروی[1]*، عبدالرسول میرقدری[2]، محمد عبداللهی ازگمی[3] و سید احمد موسوی[4]

[1و2]دانشکده فناوری اطلاعات و ارتباطات، دانشگاه جامع امام حسین (ع)، تهران، ایران

[3]دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران

[4]دانشکده ریاضی و کامپیوتر، دانشگاه شهید باهنر کرمان، کرمان، ایران

چکیده

مصالحهٔ حافظه‌زمان یک الگوریتم احتمالاتی برای وارون‌کردن توابع یک‌طرفه، با استفاده از داده‌های از پیش محاسبه‌شده است. هلمن در سال ۱۹۸۰ این روش را معرفی کرد و یک کران پایین برای احتمال موفقیت آن به‌دست آورد. پس از آن نیز تحلیل‌های پژوهش‌گران برای بررسی احتمال موفقیت، بر اساس همین کران پایین بوده است. در این مقاله، ابتدا به بررسی امید ریاضی نرخ پوشش یک ماتریس هلمن می‌پردازیم؛ سپس، این نرخ را برای ماتریس‌های هلمنی که فقط از یک زنجیره تشکیل شده‌اند، محاسبه کرده‌ایم. نشان داده‌ایم که امید ریاضی نرخ پوشش برای چنین ماتریس‌هایی بیشینه و برابر با ۰/۸۵ است. در ادامه، روش‌هایی برای تخمین دقیق‌تر این نرخ ارائه، و با روش هلمن مقایسه و درنهایت، ماتریس‌های هلمنی را معرفی کرده‌ایم که فقط از یک زنجیره تشکیل شده‌اند؛ ولی طول این زنجیره مقداری ثابت نیست و تا رسیدن به نخستین تکرار ادامه می‌یابد. به‌صورت نظری و عملی نشان داده‌ایم که احتمال موفقیت برای چنین ماتریس‌هایی بیشتر از روش هلمن است.

واژگان کلیدی: مصالحهٔ حافظه‌زمان، تابع یک‌طرفه، ماتریس هلمن، امید ریاضی نرخ پوشش

Expected coverage rate for the Hellman matrices

Nasser Hossein Gharavi*[1], Abolrasoul Mirghadri[2], Mohammad Abdollahi Azgomi[3] & Sayyed Ahmad Mousavi[4]

[1,2]Faculty of Information and Communication Technology, Imam Hossein Comprehensive University, Tehran, Iran.

[3]Faculty of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

[4]Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran

Abstract

Hellman's time-memory trade-off is a probabilistic method for inverting one-way functions, using pre-computed data. Hellman introduced this method in 1980 and obtained a lower bound for the success probability of his algorithm. After that, all further analyses of researchers are based on this lower bound.

In this paper, we first studied the expected coverage rate (ECR) of the Hellman matrices, which are constructed by a single chain. We showed that the ECR of such matrices is maximum and equal to 0.85. In this process, we find out that there exists a gap between the Hellman's lower bound and experimental coverage rate of a Hellman matrix. Specifically, this gap is larger, when considering the Hellman matrices constructed with one single chain. So, we are investigated to obtain an accurate formula for the ECR of a Hellman matrix. Subsequently, we presented a new formula that estimate the ECR of a Hellman matrix more accurately than the Hellman's lower bound. We showed that the given formula is closely match experimental data.

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

In the last, we introduced a new method to construct matrices which have much more ECR than Hellman matrices. In fact, each matrix in this new method is constructed with one single chain, which is non-repeating trajectory from a random point. So, this approach result in a number of matrices that each one contains a chain with variable length. The main advantage of this method is that we have more probability of success than Hellman method, however online time and memory requirements are increased. We have also verified theory of this new method with experimental results.

Keywords: time-memory trade-off, one-way function, Hellman matrix, expected coverage rate.

۱- مقدمه

توابع یک‌طرفه[1] به‌طورتقریبی در همه جای نظریهٔ رمزنگاری وجود دارند. امنیت برنامه‌های رمزشده، سازوکارهای تعیین اعتبار و صحت اسناد، و پروتکل‌های دیگر رمزنگاشتی، وابسته به دشورای وارون‌کردن توابع یک‌طرفه هستند. توابع یک‌طرفه توابعی هستند که محاسبهٔ آنها به‌ازای یک مقدار داده‌شده ساده، ولی وارون‌کردن آنها دشوار است. منظور ما از وارون‌کردن یک تابع یک‌طرفه، فرآیندی است که برای هر نقطهٔ داده‌شده در برد تابع، یک پیش‌تصویر[2] (در صورت وجود) ارئه کند. جستجوی جامع[3] یکی از روش‌های بدیهی برای یافتن پیش‌تصویر یک نقطهٔ داده شده است، که همهٔ ورودی‌های ممکن را بررسی می‌کند؛ اما زمان مورد نیاز برای انجام یک جستجوی جامع، بیش از حد ممکن است. ساختن یک جدول مراجعه (شامل پیش‌تصویرهای همهٔ نقاط)، می‌تواند راه‌حل دیگری باشد. در این حالت، برای توابعی که روی مجموعه‌های بزرگ عمل می‌کنند، به حافظهٔ بسیار زیادی نیازمندیم. بنابراین ایجاد تعادل بین حافظه و زمان، یا به‌عبارت دیگر مصالحه حافظه و زمان، بسیار مهم خواهد بود.

مصالحهٔ حافظه‌زمان (TMTO)[4] یک الگوریتم احتمالاتی برای وارون‌کردن توابع یک‌طرفه است. نخستین الگوریتم TMTO توسط *هلمن* در سال ۱۹۸۰ معرفی شده است [7]. این الگوریتم یک مصالحه بین زمان و حافظه پیشنهاد می‌کند که با استفاده از ذخیرهٔ مقداری از داده‌های پیش‌محاسبه‌شده در یک حافظه قابل انجام است. TMTO دارای پیچیدگی زمانی کمتر (در مرحلهٔ برخط[5]) نسبت به جستجوی جامع و پیچیدگی حافظه کمتر نسبت به جدول مراجعه است.

در ادامه تعمیم‌ها و بهبودهای بسیاری از الگوریتم TMTO ارائه شده است، که مهم‌ترین آنها روش نقاط متمایزشده[6] [2,3] و جداول رنگین‌کمانی[7] [14] هستند. برای مطالعهٔ بیشتر در این زمینه مراجع [1, 8, 9, 12] بسیار مفید هستند.

تاکنون چندین کران تقریبی برای امید ریاضی نرخ پوشش[8] یک ماتریس هلمن توسط پژوهش‌گران به‌دست آمده است [1, 7, 8, 11]؛ ولی بین کران‌های به‌دست‌آمده و داده‌های تجربی اختلاف زیادی وجود دارد.

در ادامهٔ این مقاله و در بخش ۲، به‌اختصار مصالحه حافظه‌زمان هلمن را تشریح می‌کنیم و نشان می‌دهیم که اختلاف قابل توجهی بین کران پایین به‌دست‌آمده توسط هلمن برای امید ریاضی نرخ پوشش، و داده‌های تجربی وجود دارد. در بخش سوم، امید ریاضی نرخ پوشش برای ماتریس‌های هلمن تشکیل‌شده از یک زنجیره را بررسی می‌کنیم و نشان می‌دهیم که، بیشینهٔ امید ریاضی نرخ پوشش را در این حالت داریم. این موضوع نیز اختلاف بین کران پایین به‌دست‌آمده توسط هلمن و امید ریاضی نرخ پوشش را تأیید می‌کند. در بخش ۴، امید ریاضی نرخ پوشش را برای یک ماتریس هلمن به‌طور دقیق‌تری به‌دست آورده و نرخ به‌دست‌آمده را با استفاده از داده‌های تجربی بررسی می‌کنیم؛ درنهایت، با توجه به این موضوع که ماتریس‌های هلمن تشکیل‌شده از یک زنجیره، دارای بیشینهٔ امید ریاضی نرخ پوشش هستند، ولی به‌دلیل بزرگ‌بودن t، مستلزم صرف زمان بسیار زیاد در مرحلهٔ برخط هستند، در بخش ۵ سعی می‌کنیم، این مشکل را مرتفع کرده و بدون تغییر امید ریاضی نرخ پوشش، زمان برخط را کاهش دهیم. لازم به ذکر است که در تمامی این بخش‌ها نتایج نظری به‌دست‌آمده با استفاده از داده‌های تجربی بررسی و تأیید شده‌اند.

۲- مصالحهٔ حافظه‌زمان هلمن

در سرتاسر این مقاله، برای تحلیل احتمالاتی مصالحه حافظه‌زمان، فرض می‌کنیم $f: X \to X$ یک تابع تصادفی باشد که در آن X یک فضای جستجو از اندازهٔ ثابت N است.

[1] One-way functions
[2] Pre-image
[3] Exhaustive search
[4] Time memory trade-off
[5] Online phase
[6] Distinguished points
[7] Rainbow tables
[8] Expected coverage rate (ECR)

تابع تصادفی، تابعی است که به‌طور یکنواخت و تصادفی از مجموعه همه توابع از X به X انتخاب شده باشد [10]. ترکیب k بار f را با نماد f^k نمایش خواهیم داد. همچنین یک زنجیرهٔ هلمن[1]، برای نقطهٔ داده‌شده $x\in X$ را به‌شکل زیر تعریف می‌کنیم:

$$x, f(x), f^2(x), f^3(x), \ldots$$

فرض کنید تصویر y در برد تابع f داده شده باشد. هدف الگوریتم TMTO به‌دست‌آوردن ورودی x در دامنهٔ تابع است؛ به‌طوری‌که $y=f(x)$. در روش TMTO هلمن، پارامترهای صحیح و مثبت m و t را که در *شرط توقف ماتریس*[2] $mt^2=N$ صدق می‌کنند، ثابت در نظر می‌گیریم [7]. این روش از دو مرحله تشکیل شده است: *مرحلهٔ پیش‌محاسبه*[3] و *مرحلهٔ برخط*[4].

در مرحلهٔ پیش‌محاسبه، m نقطهٔ تصادفی $x_1, x_2, \ldots, x_m$، که هر کدام مستقل از دیگر نقاط است، از مجموعه X انتخاب می‌شوند. در این صورت، ماتریس $m\times(t+1)$ هلمن به‌شکل زیر تعریف می‌شود:

$$H=\begin{pmatrix} x_1 & f(x_1) & f^2(x_1) & \cdots & f^t(x_1) \\ x_2 & f(x_2) & f^2(x_2) & \cdots & f^t(x_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_m & f(x_m) & f^2(x_m) & \cdots & f^t(x_m) \end{pmatrix}.$$

ستون‌های ابتدا و انتهای ماتریس هلمن H را در کنار یکدیگر یک *جدول هلمن*[5] می‌نامند. در مرحلهٔ پیش‌محاسبه، تنها این دو ستون ذخیره خواهند شد [7, 8].

برای به‌دست‌آوردن پیش‌تصویر یک نقطهٔ داده‌شده y در مرحلهٔ برخط، باید بررسی کنیم که آیا نقطهٔ داده‌شده در بین درآیه‌های ماتریس هلمن قرار دارد یا نه؛ چون فقط ستون‌های ابتدایی و انتهایی ماتریس هلمن ذخیره شده‌اند، یک زنجیره به شکل $y, f(y), f^2(y), \ldots$ تشکیل می‌دهیم و در هر مرحله، مقدار به‌دست‌آمده $f^i(y)$ را با نقاط انتهایی جدول هلمن مقایسه می‌کنیم. اگر $f^i(y)$ با یکی از نقاط انتهایی جدول هلمن مطابقت داشته باشد، با استفاده از نقطهٔ ابتداییِ متناظر می‌توانیم کل زنجیره را بسازیم و پیش‌تصویر مورد نظر را بیابیم. برای اطلاعات بیشتر در این زمینه مراجع [7, 8, 12] را ببینید.

فرض کنید $X_0=\{x_1, x_2, \ldots, x_m\}$ مجموعه m نقطه شروع تصادفی (که ممکن ا ست، برخی از آنها تکراری باشند) و

$$X_j \equiv f^j(X_0) \equiv \{f^j(x_0), f^j(x_1), \ldots, f^j(x_m)\}$$

مجموعه نقاط ستون $(j+1)$ ماتریس هلمن باشد. همچنین فرض کنید که H_k مجموعه تمامی درآیه‌هایی باشد که در ستون‌های نخست تا $k+1$ام ماتریس H قرار دارند؛ یعنی:

$$H_k = X_0 \cup X_1 \cup \cdots \cup X_k = \{f^j(x_i) \ : \ 1\le i\le m,\ 0\le j\le k\}.$$

متغیر تصادفی $h_{t-1}=|H_{t-1}|$ را تعداد اعضای مجموعه H_{t-1} تعریف می‌کنیم. در این صورت، از آنجایی که تنها درآیه‌های t ستون ابتدای ماتریس هلمن می‌توانند پیش‌تصویر مورد نظر ما باشند؛ لذا امید ریاضی برای متغیر تصادفی h_{t-1} بسیار مهم خواهد بود. توجه کنید که، به‌دلیل تکراری‌بودن بعضی از درآیه‌های ماتریس هلمن، $h_{t-1}\le mt$ و لذا امید ریاضی نرخ پوشش یک ماتریس هلمن به‌شکل زیر تعریف می‌شود [8]:

$$ECR(N,m,t)=\frac{1}{mt}\mathbb{E}(h_{t-1}), \qquad (۱)$$

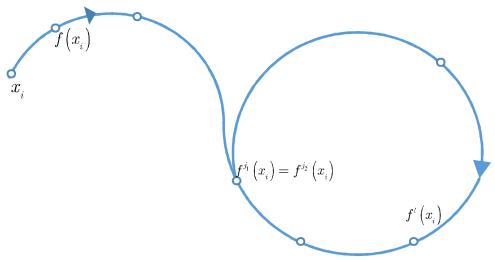
که یک اندازه برای محاسبه نرخ پوشش مورد انتظار ما برای یک ماتریس هلمن است. قابل ذکر است که امید ریاضی نرخ پوشش به‌صورت میانگین مقادیر $\frac{h_{t-1}}{mt}$، برای همه توابع f روی X و همه نقاط شروع در مجموعه X_0 محاسبه می‌شود.

ما در مصالحهٔ هلمن توقع داریم که درآیه‌های ماتریس هلمن H نقاط متمایز بسیاری از فضای X را شامل شود؛ اما بعضی محدودیت‌ها باعث می‌شوند که بسیاری از درآیه‌های ماتریس هلمن تکراری باشند. این مشکلات به‌شرح زیر هستند:

نخست این‌که، ممکن است، یک زنجیرهٔ هلمن (یا یک سطر ماتریس هلمن) با خودش ادغام شود و در دور بیفتد؛ این

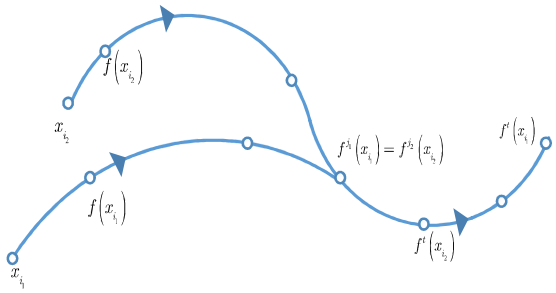
[1] Hellman chain
[2] Matrix stopping rule
[3] Precomputation phase
[4] Online phase
[5] Hellman table

وضعیت زمانی اتفاق می‌افتد که برای $0 \le j_1 < j_2 \le t$ در سطر iام ماتریس هلمن داشته باشیم $f^{j_1}(x_i) = f^{j_2}(x_i)$ (شکل (۱) را ببینید).



(شکل-۱): دور در یک زنجیره هلمن
(Figure-1): Loop in a Hellman chain

دوم این‌که، ممکن است دو یا چند زنجیرهٔ هلمن با یکدیگر ادغام شوند. در این حالت برای دو نقطه شروع متمایز مثل x_{i_1} و x_{i_2} و دو مقدار صحیح j_1 و j_2 خواهیم داشت $f^{j_1}(x_{i_1}) = f^{j_2}(x_{i_2})$. یعنی در این مکان برخورد بین دو زنجیره اتفاق می‌افتد و لذا در ادامه نیز هر دو زنجیره با یکدیگر ادغام می‌شوند (شکل (۲) را ببینید).



(شکل-۲): ادغام دو زنجیره هلمن
(Figure-2): Merge of two Hellman chain

لذا پیشنهاد شده است که علاوه بر کاهش تعداد سطرهای یک ماتریس هلمن، r ماتریس هلمن با استفاده از توابع مرکب $f_i = R_i \circ f$، برای $1 \le i \le r$، ایجاد شوند. درحقیقت، هر $R_i : X \to X$ یک تابع دوسویی ساده (از دیدگاه محاسباتی) است، که آن را یک تابع احیا[1] می‌نامند. به‌عنوان مثال، این تابع ساده می‌تواند یک تابع جایگشت و یا جمع به پیمانه N با یک مقدار ثابت باشد. برای اطلاعات بیشتر در این زمینه مرجع [8] را ببینید. هنگامی که با استفاده از r تابع احیای مستقل، r ماتریس هلمن $m \times (t+1)$ را مورد استفاده قرار داده باشیم، احتمال موفقیت[2] مصالحهٔ هلمن به‌شکل زیر محاسبه می‌شود [8]:

$$1-\left(1-ECR(N,m,t).\frac{mt}{N}\right)^r.$$

هنگامی که $N \to \infty$ این مقدار به‌طورتقریبی برابر است با:

$$1-e^{-ECR(N,m,t)\frac{mt}{N}r}. \qquad (۲)$$

بنابراین برای محاسبهٔ احتمال موفقیت مصالحه هلمن، کافی است، امید ریاضی نرخ پوشش را برای یکی از ماتریس‌های هلمن به‌دست آوریم؛ پس تمرکز ما در ادامهٔ این مقاله بر روی امید ریاضی نرخ پوشش یک ماتریس هلمن خواهد بود.

هلمن در [7]، کران پایین زیر را برای امید ریاضی نرخ پوشش یک ماتریس هلمن به‌دست آورد:

$$ECR(N,m,t) \ge \frac{1}{mt}\sum_{i=1}^{m}\sum_{j=1}^{t}\left(1-\frac{it}{N}\right)^j. \qquad (۳)$$

هلمن به‌صورت عددی تشریح کرد، هنگامی که $mt^2 = N$ مقدار این کران پایین به‌طورتقریبی برابر با ۰/۸ است. در مرجع [11]، قسمت سمت راست معادله (۳) با دقت بیشتر محاسبه شد و کران زیر در حالت $mt^2 = N$ و $m,t \gg 1$ به‌دست آمد:

$$ECR(N,m,t) \ge \int_0^1 \frac{1-e^{-x}}{x}\,dx \approx 0.796599.$$

فرض کنید $r \approx t$ ماتریس هلمن با استفاده از توابع احیای مستقل از هم ساخته شده باشند. حتی فرض کنید تعداد نقاط تکراری در یک ماتریس هلمن بسیار کم باشد؛ یعنی $ECR(N,m,t) \approx 1$. در این صورت، با درنظرگرفتن شرط توقف ماتریس $mt^2 = N$ و با استفاده از معادله (۲)، احتمال موفقیت مصالحه هلمن بیشینه $1-e^{-1} \approx 0.632$ خواهد بود، که احتمال موفقیت خوبی نیست.

ما در شکل (۳) با استفاده از معادله (۳) و شرط توقف ماتریس $mt^2 = N$، منحنی کران پایین داده شده توسط هلمن را با استفاده از نرم‌افزار متلب[3] رسم کرده‌ایم. منحنی آبی‌رنگ در این شکل، کران را بر حسب لگاریتم طول زنجیره (یعنی $\log_2(t)$) برای حالت $N = 2^{16}$ نمایش می‌دهد. هر چند در مورد اعتبار این کران شکی وجود ندارد، اما در ادامه نشان خواهیم داد که اختلاف به‌نسبه زیادی با داده‌های تجربی وجود دارد. درحقیقت، طبق شرط توقف ماتریس، برای tهای بزرگ و نزدیک به $\sqrt{N}$، m عددی متناهی خواهد بود؛ لذا شرط $m \gg 1$ در کران به‌دست‌آمده توسط هلمن برآورده نمی‌شود و بنابراین نمی‌توان از این کران استفاده کرد.

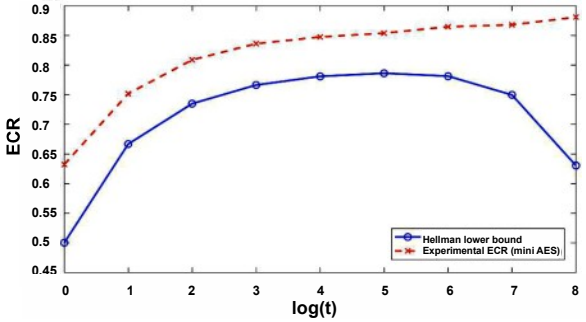
[1] Reduction function
[2] Success probability
[3] MATLAB

در این مقاله، برای تولید داده‌های تجربی از الگوریتم رمز mini AES [13] استفاده شده است. اهداف ما برای استفاده از این الگوریتم، کوچک‌بودن فضای کلید نسبت به الگوریتم رمز AES و استانداردبودن آن است. فرض کنید $p_0 \in \{0,1\}^{16}$ یک متن تصادفی و ثابت باشد. تابع $f:\{0,1\}^{16} \to \{0,1\}^{16}$ را به‌صورت:

$$f(k) = \text{mini AES}(p_0, k)$$

تعریف می‌کنیم. این تابع همه خواص یک تابع تصادفی را دارد. برای دیدن برخی از این خواص مرجع [4] را ببینید.

در ادامه نقطهٔ شروع تصادفی را برای مقادیر مختلف t، که $0 \le \log_2(t) \le 8$، تعداد $m = 2^{16}/t^2$ انتخاب کرده‌ایم؛ سپس برای هر یک از این مقادیر یک ماتریس هلمن H با استفاده تابع تصادفی f و نرم‌افزار متلب ایجاد کرده‌ایم. درنهایت، به‌صراحت نرخ پوشش h_{t-1}/mt را محاسبه کرده‌ایم. اگر این کار را، برای مقادیر مختلف t، به تعداد بسیار زیاد انجام دهیم و میانگین مقادیر به‌دست‌آمده را رسم کنیم، منحنی قرمزرنگ در شکل (۳) به‌دست می‌آید. این نمودار میانگین نرخ پوشش آزمایش‌های تجربی ما را بر حسب لگاریتم طول زنجیره، یعنی $\log_2(t)$، نشان می‌دهد.

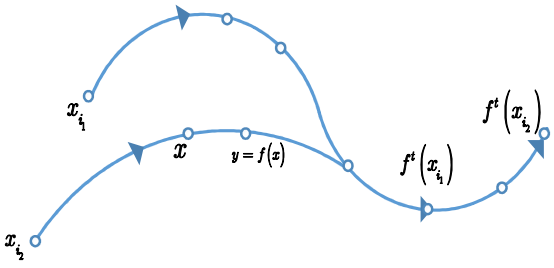


(شکل-۳): منحنی کران پایین داده‌شده توسط هلمن در معادله (۳) بر حسب $\log_2(t)$ برای $N = 2^{16}$ و $m = N/t^2$ (منحنی با علامت دایره)؛ منحنی تجربی میانگین نرخ پوشش ماتریس‌های هلمن ساخته‌شده توسط الگوریتم mini AES بر حسب $\log_2(t)$ (منحنی خط‌چین با علامت ضرب).

(Figure-3): Given lower bound by Hellman in Equation (3) in logarithmic scale for $N = 2^{16}$ and $m = N/t^2$ (curve with circle markers); Experimental average of coverage rate of Hellman matrices constructed by mini AES algorithm in logarithmic scale (curve with cross markers).

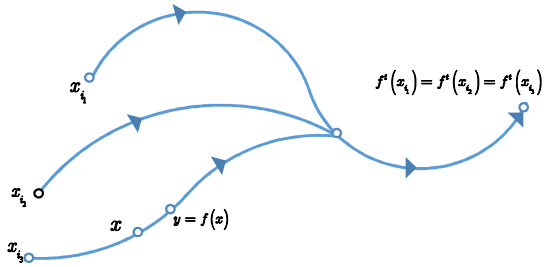
همان‌طور که در شکل (۳) مشاهده می‌کنید، منحنی کران ارائه‌شده توسط هلمن پایین‌تر از منحنی تجربی قرار گرفته و واقعاً یک کران پایین است؛ ولی شکاف بین منحنی نظری و تجربی (به‌خصوص در حالت $t = \sqrt{2^{16}} = 2^8$ یا $m = 1$) مشهود است. از طرف دیگر منحنی تجربی نشان می‌دهد که امید ریاضی نرخ پوشش، صعودی است و برای tهای بزرگ و نزدیک به $\sqrt{N}$ افزایش می‌یابد. البته ما این موارد را به‌صورت دقیق‌تر در لم ۱-۳ و نتیجه ۲-۳ بخش سوم نشان خواهیم داد.

یکی از مشکلات بزرگی که در مرحلهٔ برخط و به‌علت دور و ادغام در زنجیره‌های هلمن گریبان‌گیر ما خواهد شد، پدیده *هشدار خطا*[1] است. برای نمونه شکل‌های (۴و۵) را ببینید. در شکل (۴) نقاط انتهایی هر دو زنجیره، یعنی $f^t(x_{i_1})$ و $f^t(x_{i_2})$ در قسمت ادغام دو زنجیره قرار دارند. در این صورت، اگر در مرحلهٔ برخط تصویر y روی زنجیرهٔ دوم داده شده باشد، با شروع ساخت زنجیرهٔ برخط از y، ابتدا به نقطه انتهایی $f^t(x_{i_1})$ برخورد می‌کنیم و لذا به‌اشتباه نتیجه خواهیم گرفت که پیش‌تصویر y روی زنجیرهٔ نخست قرار دارد. بنابراین با شروع از نقطه x_{i_1} نمی‌توانیم به پیش‌تصویر مورد نظر برسیم. حالت دیگر هشدار خطا در شکل (۵) نمایش داده شده است. همان‌طور که می‌بینید، ممکن است، متناظر با چند نقطهٔ شروع فقط یک نقطهٔ انتهایی داشته باشیم. در این وضعیت باید در مرحلهٔ برخط، چندین زنجیره با استفاده از این نقاط شروع تشکیل دهیم.



(شکل-۴): پدیده هشدار خطا

(Figure-4): False alarm situation



(شکل-۵): پدیده هشدار خطا

(Figure-5): False alarm situation

[1] False alarm

۳- امید ریاضی نرخ پوشش در حالت $m=1$

هدف ما در این بخش محاسبهٔ دقیق امید ریاضی نرخ پوشش برای ماتریس‌های هلمن تشکیل‌شده از یک زنجیره به‌طول t است. برای این کار به لم و نتیجه‌ای که در ادامه می‌آیند نیازمندیم. قابل ذکر است که لم و نتیجهٔ زیر، در مراجع بسیاری از جمله [4, 5, 6] نیز به شکل متفاوتی بررسی شده‌اند.

لم ۱-۳. فرض کنید $f: X \to X$ یک تابع تصادفی روی فضای جستجوی X از اندازه N باشد و

$$\rho = \left|\left\{f^j(x) : j = 0,1,2,\ldots\right\}\right|$$

پیشامد تعداد اعضای زنجیره بدون تکرار با شروع از نقطه تصادفی $x \in X$ باشد. در این صورت هنگامی که N به بی‌نهایت میل می‌کند $\Pr[\rho \geq t] \approx e^{\frac{-t^2}{2N}}$.

اثبات. توجه کنید که فرآیند ساختن یک زنجیره با استفاده از نقطهٔ شروع تصادفی x را می‌توان به شکل برداشتن مهره از یک گلدان شامل N مهره شماره‌گذاری‌شده و با جای‌گذاری مدل کرد. بنابراین:

$$\Pr[\rho \geq t] = \frac{N}{N}\frac{N-1}{N}\frac{N-2}{N}\cdots\frac{N-t+1}{N}.$$

می‌دانیم هنگامی که $i \ll N$ داریم $\frac{N-i}{N} \approx e^{-i/N}$. پس:

$$\Pr[\rho \geq t] \approx \prod_{i=0}^{t-1} e^{-i/N} \approx e^{-t^2/2N}$$

نتیجه ۲-۳. فرض کنید $f: X \to X$ تابعی تصادفی روی فضای جستجو X از اندازه N باشد و ρ پیشامد تعداد اعضای زنجیرهٔ بدون تکرار با شروع از یک نقطهٔ تصادفی در X باشد. در این صورت هنگامی که $N \to \infty$ داریم:

$$\mathbb{E}[\rho] \approx \sqrt{\frac{\pi N}{2}}.$$

اثبات. هنگامی که $N \to \infty$ داریم:

$$\mathbb{E}[\rho] = \sum_{t=0}^{N} \Pr[\rho \geq t] \approx \sum_{t=0}^{N} e^{-t^2/2N} \approx N\int_0^1 e^{-\frac{N}{2}x^2}\,\mathrm{d}x = \sqrt{\frac{\pi N}{2}}.$$

حال آماده‌ایم که با استفاده از نتایج به‌دست‌آمده، امید ریاضی نرخ پوشش برای ماتریس‌های هلمن تشکیل‌شده از یک زنجیره را (بدون شرط توقف ماتریس) به‌دست آوریم.

قضیه ۳-۳. فرض کنید $f: X \to X$ یک تابع تصادفی روی فضای جستجوی X از اندازه N باشد و H ماتریس هلمن $1 \times (t+1)$ باشد که با شروع از یک نقطهٔ تصادفی در X ایجاد شده است. همچنین فرض کنید $t = \alpha\sqrt{N}$ برای یک $\alpha \geq 0$. در این صورت، امید ریاضی نرخ پوشش H وقتی که $N \to \infty$ به‌طورتقریبی برابر است با:

$$ECR(N,1,t) \approx \frac{\sqrt{2\pi}}{\alpha}\int_0^{\alpha} \varphi(z)\,dz, \qquad (۴)$$

که در آن $\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ تابع چگالی احتمال توزیع نرمال استاندارد است.

اثبات. فرض کنید ρ پیشامد تعداد اعضای زنجیره بدون تکرار با شروع از یک نقطه تصادفی در X باشد؛ بنابراین برای $t = \alpha\sqrt{N}$ داریم:

$$ECR(N,1,t) = \frac{1}{t}\mathbb{E}(h_t) = \frac{1}{t}\left(\sum_{k=0}^{t}\Pr[\rho \geq k]\right).$$

حال با استفاده از لم ۱-۳ به‌دست می‌آوریم:

$$ECR(N,1,t) \approx \frac{1}{t}\sum_{k=0}^{t} e^{\frac{-k^2}{2N}}.$$

درنتیجه هنگامی که $N \to \infty$ داریم:

$$\begin{aligned} ECR(N,1,t) &\sim \int_0^1 e^{-\frac{t^2}{2N}x^2}\,\mathrm{d}x \\ &= \sqrt{2\pi}\,\frac{\sqrt{N}}{t}\int_0^{\frac{t}{\sqrt{N}}} \varphi(z)\,\mathrm{d}z \\ &= \sqrt{2\pi}\,\frac{1}{\alpha}\int_0^{\alpha} \varphi(z)\,\mathrm{d}z \end{aligned}$$

که در آن $\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ تابع چگالی احتمال توزیع نرمال استاندارد است.

نتیجه ۴-۳. امید ریاضی نرخ پوشش برای ماتریس هلمن تشکیل‌شده از یک زنجیره (به همراه شرط توقف ماتریس $mt^2 = N$)، برابر با ۰٫۸۵ است.

اثبات. توجه کنید که در حالت $m=1$ با استفاده از شرط توقف ماتریس، $t = \sqrt{N}$ است و لذا $\alpha = 1$ خواهد بود؛ بنابراین با استفاده از قضیه ۳-۳ داریم:

$$\lim_{t\to\infty} ECR(N,1,t) \approx \sqrt{2\pi}\cdot 0.3413 = 0.85.$$

همان‌گونه‌که در این نتیجه دیدیم، یک ماتریس هلمن با پارامترهای $m=1$ و $t = \sqrt{N}$ دارای امید ریاضی نرخ پوشش بسیار خوبی است؛ اما مشکل اصلی این‌گونه ماتریس‌ها، صرف

زمان بسیار زیاد (از مرتبه $\sqrt{N}$) در مرحلهٔ برخط است، که باعث می‌شود، کاربرد عملی چندانی نداشته باشند. در بخش ۵ سعی می‌کنیم این مشکل را مرتفع کرده و علاوه‌بر عدم تغییر نرخ پوشش، زمان برخط را کاهش دهیم. از طرف دیگر، اگر نرخ به‌دست‌آمده را در این نتیجه با شکل (۳) در حالت $\log(t)=8$ مقایسه کنید، اختلاف با کران پایین به‌دست‌آمده توسط هلمن و مطابقت آن با نتایج تجربی مشخص خواهد شد. درحقیقت، در حالتی $m=1$ نمی‌توانیم از فرمول هلمن برای امید ریاضی نرخ پوشش استفاده کنیم؛ لذا، ما در بخش ۴ به بررسی دقیق‌تر امید ریاضی نرخ پوشش یک ماتریس هلمن خواهیم پرداخت.

۴- محاسبهٔ امید ریاضی نرخ پوشش یک ماتریس هلمن

هدف ما در این بخش، محاسبه دقیق‌تر امید ریاضی نرخ پوشش برای یک ماتریس هلمن خواهد بود. برای این کار، ابتدا به لم زیر نیازمندیم که اثبات آن را در کتاب‌های استاندارد نظریه احتمال می‌توان یافت.

لم ۱-۴. فرض کنید گلدانی شامل N توپ که به‌طور مجزا شماره‌گذاری شده‌اند، داشته باشیم؛ همچنین فرض کنید که m توپ را یکی‌یکی و با جای‌گذاری از گلدان برمی‌داریم؛ در این صورت انتظار داریم که $N\left(1-\left(1-\frac{1}{N}\right)^m\right)$ توپ متمایز ببینیم؛ این امید ریاضی وقتی $N\to\infty$ به‌طورتقریبی برابر با $N\left(1-e^{-m/N}\right)$ است.

امید ریاضی پوشش یک تابع تصادفی f به شکل امید ریاضی تعداد نقاط موجود در برد تابع یا $\mathbb{E}\left(\left|f(X)\right|\right)$ تعریف می‌شود. توجه کنید که، با قراردادن $m=N$ در لم ۱-۴ می‌توان امید ریاضی پوشش یک تابع تصادفی را به‌سادگی به‌دست آورد. برای دیدن اثبات‌های دیگر از نتیجه زیر، قضیهٔ دو در مرجع [4] را ببینید.

نتیجه ۲-۴. امید ریاضی پوشش یک تابع تصادفی $f:X\to X$ به‌طورتقریبی برابر است با:

$$\mathbb{E}\left(\left|f(X)\right|\right)\sim N\left(1-e^{-1}\right).$$

در مدل [7] برای محاسبهٔ امید ریاضی نرخ پوشش، فقط از فرض تصادفی‌بودن تابع f استفاده شده و شرط تابع‌بودن f در نظر گرفته نشده است. در ادامه قصد داریم با استفاده از لم ۱-۴ تعداد عناصر متمایز در هر ستون را به‌ترتیب شمارش کرده و با استفاده از نقش تابع f در ستون بعد، امید ریاضی نرخ پوشش (بدون شرط توقف ماتریس) را به‌دست آوریم (بخش ضمیمه مرجع [14] را ببینید).

حدس ۳-۴. فرض کنید $f:X\to X$ یک تابع تصادفی روی فضای جستجو از اندازه ثابت N باشد؛ همچنین فرض کنید که H ماتریس هلمن $m\times(t+1)$ باشد که با استفاده از m نقطه تصادفی در X ایجاد شده است؛ در این صورت امید ریاضی نرخ پوشش H به‌طورتقریبی برابر است با:

$$ECR(N,m,t)\approx\frac{N}{mt}s_t \qquad (۵)$$

که در آن برای $\alpha=\sqrt{\frac{m}{2N}}$ داریم:

$$s_k\approx 2\alpha\left(1-\alpha^2\right)\frac{\tanh(\alpha k)}{1+\alpha\tanh(\alpha k)}. \qquad (۶)$$

طرح اثبات. توجه نمایید که:

$$\mathbb{E}\left(h_{t-1}\right)=\mathbb{E}\left(\left|H_{t-1}\right|\right)=m_0+m_1+\cdots+m_{t-1}$$

که در آن $m_0=\mathbb{E}\left(\left|H_0\right|\right)$ و $m_j=\mathbb{E}\left(\left|H_j\setminus H_{j-1}\right|\right)$ تعداد اعضای متمایز ستون $j+1$ است که در ستون‌های j و ما قبل آن موجود نیستند.

با به کاربردن لم ۱-۴ برای ستون نخست ماتریس هلمن، امید ریاضی تعداد نقاط متمایز در ستون نخست برابر است با:

$$m_0=\mathbb{E}\left(\left|H_0\right|\right)=N\left(1-e^{-m/N}\right).$$

توجه کنید که اگر $x_i\in H_0$ برای چند i یکسان باشند، آن‌گاه $f(x_i)\in H_1$ها نیز یکسان خواهند بود و بنابراین تنها درآیه‌هایی مثل x_i متعلق به ستون نخست ماتریس هلمن که متعلق به مجموعه H_0 هستند، می‌توانند درایه‌های جدیدی در ستون دوم ایجاد کنند (اینجا است که نقش تابع‌بودن f تأثیر دارد). حال فرض کنید که در ستون نخست m_0 عضو متمایز و ثابت وجود دارد (توجه کنید که این کار از نظر احتمالاتی دقیق نیست. تبصره ۴-۴ را ببینید). نقش تصادفی تابع f در ایجاد ستون دوم ایجاب می‌کند که به آن مثل برداشتن m_0 توپ از یک گلدان شامل N توپ برچسب‌دار و متمایز با جایگذاری نگاه کنیم. بنابراین لم ۱-۴ می‌گوید، انتظار داریم $N\left(1-e^{-m_0/N}\right)$ درآیه متمایز در ستون دوم داشته باشیم. پس تعداد درآیه‌های متمایز در ستون دوم که در ستون نخست نیز نباشند، برابر است با:

$$m_1 = \mathbb{E}\left(\left|H_1 \setminus H_0\right|\right) = \left(1-\tfrac{m_0}{N}\right) N\left(1-e^{-\frac{m_0}{N}}\right).$$

به همین شکل، اگر $f^{j-1}\left(x_i\right) \in H_{j-1}$ آن‌گاه $f^j\left(x_i\right) \in H_j$ و بنابراین تنها درایه‌هایی مثل $f^j\left(x_i\right)$ متعلق به ستون $(j+1)$ام ماتریس هلمن، که متعلق به مجموعه $H_j \setminus H_{j-1}$ هستند، می‌توانند درایه‌های جدیدی در ستون $j+2$ ایجاد کنند. بنابراین انتظار داریم $N\left(1-e^{m_j/N}\right)$ درایه متمایز در ستون $j+2$ داشته باشیم؛ به‌طوری‌که در ستون‌های ماقبل عضویت ندارند. مثل قبل فرض کنید که در $H_j \setminus H_{j-1}$ به تعداد m_j عضو متمایز و ثابت داریم. درنتیجه:

$$m_{j+1} = \mathbb{E}\left(\left|H_{j+1} \setminus H_j\right|\right) = \left(1-\sum_{i=0}^{k}\tfrac{m_i}{N}\right) N\left(1-e^{-\frac{m_j}{N}}\right),$$

حال با استفاده از نمادگذاری $s_j = \sum_{i=0}^{j-1}\frac{m_i}{N}$ به‌دست می‌آوریم:

$$1-s_{k+1} = \left(1-s_1\right)e^{-s_k}, \qquad (k \geq 1).$$

فرض کنید $c = e^{-m/N}$. در این صورت با استفاده از شرط اولیه $s_1 = \frac{m_0}{N} = 1-c$ خواهیم داشت:

$$s_0 = 0, \qquad s_{k+1} = 1-ce^{-s_k}.$$

سرانجام با استفاده از معادله (۱) امید ریاضی نرخ پوشش برای یک ماتریس هلمن به شکل زیر است:

$$ECR\left(N,m,t\right) = \tfrac{1}{mt}\mathbb{E}\left(h_{t-1}\right) = \frac{m_0+\cdots+m_{t-1}}{mt} = \tfrac{N}{mt}s_t,$$

برای هر $t \geq 1$.

در انتها، برای به‌دست‌آوردن یک شکل بسته برای فرمول بازگشتی، داریم:

$$\frac{\mathrm{d}\,s_k}{\mathrm{d}\,k} \approx s_{k+1}-s_k = \left(1-s_k\right)-e^{-\left(s_k+\frac{m}{N}\right)}$$

لذا با استفاده از تقریب بسط تیلور تابع نمایی $e^x \approx 1+x+\frac{1}{2}x^2$ معادله دیفرانسیل زیر به‌دست می‌آید:

$$\frac{\mathrm{d}\,s_k}{\mathrm{d}\,k} \sim \frac{m}{N}-\frac{1}{2}\left(s_k+\frac{m}{N}\right)^2, \qquad s_0 = 0.$$

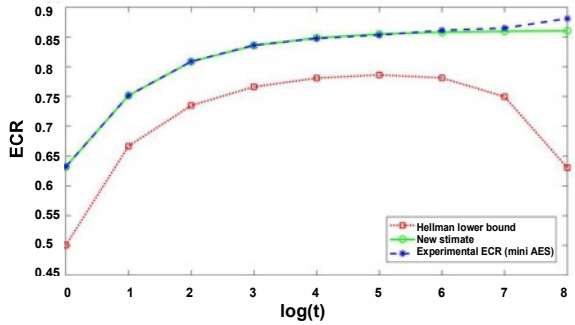
با حل این معادله دیفرانسیل داریم:

$$k \approx \alpha^{-1}\left(\tanh^{-1}\left(\tfrac{1}{2}\alpha^{-1}s_k+\alpha\right)-\tanh^{-1}\left(\alpha\right)\right).$$

که در آن $\alpha = \sqrt{\frac{m}{2N}}$. با حل معادله بالا، معادله (۶) به‌سادگی به‌دست می‌آید.

تبصره ۴-۴. توجه کنید که در طرح اثبات حدس ۳-۴ فرض کردیم که تعداد نقاط متمایز در ستون نخست، برابر با امید ریاضی تعداد نقاط متمایز در این ستون است؛ سپس به همین تعداد نقطه از گلدان شامل N توپ برچسب‌دار و متمایز با جایگذاری انتخاب کردیم. این فرض برای ستون‌های بعد نیز در نظر گرفته شده است. هر چند ممکن است، این تقریب از نقطه نظر احتمالاتی دقیق نباشد؛ اما منحنی‌های تجربی در ادامه این بخش نشان خواهند داد که این تقریب با منحنی تجربی به‌دست‌آمده توسط الگوریتم mini AES مطابقت بسیار خوبی دارد؛ لذا تقریب دقیق‌تر از امید ریاضی نرخ پوشش یک ماتریس هلمن، همچنان می‌تواند به‌عنوان یک مسأله باز پژوهشی مطرح باشد.

منحنی خط‌چین (با علامت ستاره) در شکل (۶)، امید ریاضی نرخ پوشش به‌دست‌آمده در حدس ۳-۴ را با فرض $mt^2 = N = 2^{16}$، نمایش می‌دهد. همان‌گونه‌که در این شکل مشاهده می‌کنید، منحنی خط‌چین، بسیار نزدیک به منحنی تجربی به‌دست‌آمده توسط الگوریتم mini AES (منحنی با علامت دایره) است. بنابراین تخمین ارائه‌شده در حدس ۳-۴ بسیار دقیق‌تر از تخمین ارائه‌شده توسط هلمن (منحنی نقطه‌چین با علامت مربع) است.



(شکل-۶): منحنی امید ریاضی نرخ پوشش ارائه شده در حدس ۳-۴ بر حسب $\log_2(t)$ برای $N = 2^{16}$ (منحنی خط‌چین با علامت ستاره)؛ منحنی نرخ پوشش تجربی با استفاده از الگوریتم mini AES (منحنی با علامت دایره)؛ کران پایین ارائه شده توسط هلمن (منحنی نقطه‌چین با علامت مربع).

(Figure-6): Expected coverage rate given in Conjecture 4-3 in logarithmic scale for $N = 2^{16}$ (dashed curve with star markers); Experimental coverage rate obtained by using mini AES algorithm (curve with circle markers); Hellman lower coverage (dotted curve with square markers).

نکته قابل توجه در شکل (۶) این است که، منحنی امید ریاضی نرخ پوشش ارائه‌شده در حدس ۳-۴ در نقطه $t = \sqrt{N}$ نادقیق است. درحقیقت، همان‌طورکه در این شکل می‌بینید، منحنی تجربی (با علامت دایره) در این نقطه بالاتر

از منحنی خط‌چین (با علامت ستاره) قرار گرفته است. البته این موضوع دور از انتظار ما نیست؛ زیرا روش به‌دست‌آوردن امید ریاضی نرخ پوشش در طرح اثبات حدس ۳-۴ به‌گونه‌ای است که برای ماتریس‌های هلمن تشکیل‌شده از یک زنجیره ($m=1$) نادقیق است. توجه کنید که، ما این حالت را درقبل در بخش سوم بررسی کرده‌ایم.

۵- روشی جدید برای تولید ماتریس هلمن

همان‌طورکه در بخش چهارم مشاهده کردید، امید ریاضی نرخ پوشش یک ماتریس هلمن، با شرط توقف ماتریس، یک تابع صعودی بر اساس t است. بنابراین در حالت $m=1$ و $t=\sqrt{N}$ بیشترین امید ریاضی نرخ پوشش را خواهیم داشت. در بخش سوم، این امید ریاضی را به‌دقّت محاسبه کردیم و مقدار ۰/۸۵ را به‌دست آوردیم؛ اما گفتیم مهم‌ترین مشکلی که باعث می‌شود، این حالت از نظر عملی ناکارآمد باشد، مدت‌زمانی (از مرتبه $\sqrt{N}$) است که در مرحلهٔ برخط باید صرف شود تا به پیش‌تصویر نقطهٔ داده‌شده برسیم. در این بخش روشی جدید برای تولید یک ماتریس هلمن پیشنهاد می‌کنیم که دارای بیشترین امید ریاضی نرخ پوشش است. درحقیقت، فرآیند زیر را دنبال خواهیم کرد.

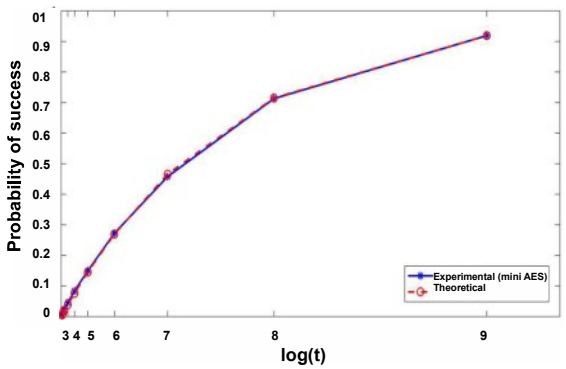
فرض کنید تابع تصادفی f داده شده باشد، هدف ما در این بخش ساختن r ماتریس است که در هر یک از آنها از توابع احیای متمایز و مستقل استفاده شده است. در هر ماتریس تنها یک زنجیره به‌طول غیر ثابت داریم. درحقیقت، این زنجیره از یک نقطه تصادفی شروع و تا قبل از رسیدن به نخستین تکرار ادامه می‌یابد. توجه کنید که در این روش، زمان مورد استفاده برای مرحلهٔ پیش‌محاسبه افزایش می‌یابد؛ زیرا در طی ساخت یک زنجیره، باید بتوانیم دور را تشخیص داده و قبل از رسیدن به نخستین تکرار، ساخت زنجیره را پایان دهیم. یکی از مزایای این روش این است که نرخ پوشش هر یک از این ماتریس‌ها به‌طوردقیق برابر با یک خواهد بود؛ یعنی هیچ نقطهٔ تکراری در این گونه ماتریس‌ها وجود ندارد. از طرف دیگر، در مرحلهٔ برخط، پدیده هشدار خطا نیز رخ نخواهد داد؛ زیرا به‌وضوح ادغام در این‌گونه ماتریس‌ها رخ نمی‌دهد.

همان‌طورکه در لم ۱-۳ و نتیجه آن مشاهده کردیم، امید ریاضی تعداد اعضای یک زنجیرهٔ تصادفی برابر با $\sqrt{\pi N/2}$ است. حال، چون $m=1$ و $t\sim\sqrt{\pi N/2}$ است، با استفاده از معادله (۲)، احتمال موفقیت روش ما برای r ماتریس به‌طورتقریبی برابر است با:

$$1-e^{-r\sqrt{\frac{\pi}{2N}}} \qquad (۷)$$

منحنی این فرمول بر حسب $\log_2(r)$ برای $N=2^{16}$ در شکل (۷) با منحنی خط‌چین (با علامت دایره) رسم شده است.

برای مقایسه این منحنی با داده‌های تجربی، مشابه با قبل، از الگوریتم mini AES استفاده می‌کنیم. برای این کار تعداد r، $0\le\log(r)\le 9$، ماتریس به روش جدید، با استفاده از توابع $f_i(x)=f(x\oplus i)$، که در آن $\oplus$ تابع XOR است، تشکیل می‌دهیم؛ سپس تعداد نقاط متمایز در کل جداول را به‌صراحت محاسبه می‌کنیم. با تقسیم عدد به‌دست‌آمده بر 2^{16} موفقیت این r ماتریس هلمن به‌دقّت به‌دست می‌آید. اگر این کار را به دفعات زیاد انجام داده و میانگین بگیریم، منحنی با علامت ستاره در شکل (۷) حاصل می‌شود. همان‌طورکه در این شکل دیده می‌شود، منحنی‌های نظری و تجربی مطابقت بسیار خوبی با یکدیگر دارند.



(شکل-۷): منحنی امید ریاضی نرخ پوشش r ماتریس به روش پیشنهادی بر حسب $\log_2(r)$ به صورت نظری (منحنی با علامت دایره) در مقایسه با منحنی نرخ پوشش r ماتریس به روش پیشنهادی با استفاده از الگوریتم mini AES (منحنی خط‌چین با علامت ستاره).

(Figure-7): Theoretical expected coverage rate of r matrix with presented method in logarithmic scale (curve with circle markers); Experimental coverage rate of r matrix with presented method in logarithmic scale using mini AES algorithm (dashed curve with star markers).

یکی از مشکلاتی که باعث می‌شود r ماتریس تشکیل‌شده به روش پیشنهادی از نظر عملی ناکارآمد باشند، این است که در مرحلهٔ برخط باید زمان بسیار زیادی، از مرتبهٔ $rt\approx r\sqrt{\pi N/2}$ را صرف کنیم تا به پیش‌تصویر مورد نظر برسیم. برای این‌که این مشکل را تا حدی جبران کنیم، به‌ازای

یک t' مناسب (بر حسب زمان مورد نیاز کاربر در مرحلهٔ برخط و حافظه در دسترس او)، زنجیرهٔ هر یک ماتریس‌ها را به‌شکل زیر به زیرزنجیره‌هایی با طول t' تقسیم می‌کنیم:

$$\begin{aligned} & x_{1,0} \to x_{1,1} \to x_{1,2} \to \cdots \to x_{1,t'} \to \\ & \to x_{2,0} \to x_{2,1} \to x_{2,2} \to \cdots \to x_{2,t'} \to \\ & \quad \vdots \\ & \to x_{m',0} \to x_{m',1} \to x_{m',2} \to \cdots \to x_{m',t'} \end{aligned}$$

که در آن $m't' \sim \sqrt{\pi N/2}$. همچنین در این ماتریس $f\left(x_{i,j}\right) = x_{i,j+1}$ برای $0 \le j < t'$ و $1 \le i \le m'$، و $f\left(x_{i,t'}\right) = x_{i+1,0}$ برای $2 \le i \le m'$.

توجه کنید که با این کار نرخ پوشش و درنتیجه احتمال موفقیت تفاوتی با قبل نخواهد داشت؛ ولی حافظه مورد نیاز افزایش می‌یابد. درحقیقت، زمان برخط از مرتبه $t'r$ و حافظه مورد نیاز از مرتبه $m'r$ است.

مثال. فرض کنید $t' = N^{\alpha}$، $m' = N^{\beta}$ و $r = N^{\gamma}$. در این صورت، چون $m't' \approx \rho \approx \sqrt{N}$، پس $\alpha + \beta \approx \frac{1}{2}$. درنتیجه، زمان برخط از مرتبه $t'r \approx N^{\alpha+\gamma}$ و حافظه از مرتبه $t'r \approx N^{\beta+\gamma}$ است. از طرفی با استفاده از معادله (۷)، احتمال موفقیت به‌طورتقریبی برابر است با:

$$1 - \exp\left(-\left(N^{\gamma-\frac{1}{2}}\right)\sqrt{\tfrac{\pi}{2}}\right).$$

در جدول (۱)، مرتبه زمان برخط، حافظه و احتمال موفقیت برای چند نمونه محاسبه شده است. همان‌طورکه می‌دانیم با استفاده از روش هلمن، در بهترین حالت، که $m \approx t \approx r \approx N^{1/3}$، زمان برخط از مرتبه $tr \approx N^{2/3}$ و حافظه از مرتبه $mr \approx N^{2/3}$ است. همچنین بیشینهٔ احتمال موفقیت $1 - e^{-1} \approx 0.632$ است.

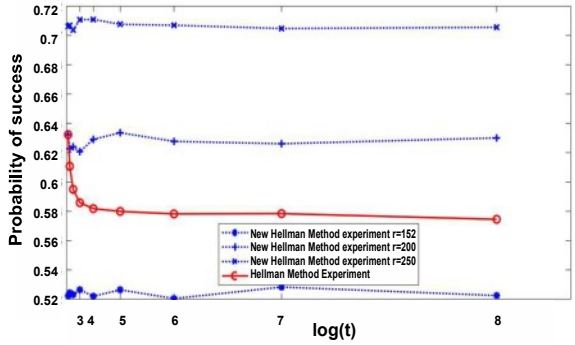
(جدول-۱): چند نمونه از پارامترها برای روش پیشنهادی

(Table-1): some parameters for presented method

α	β	γ	زمان برخط	حافظه	احتمال موفقیت
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$N^{3/4}$	$N^{3/4}$	0.714
$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$N^{5/6}$	$N^{2/3}$	0.714
$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$N^{2/3}$	$N^{5/6}$	0.714

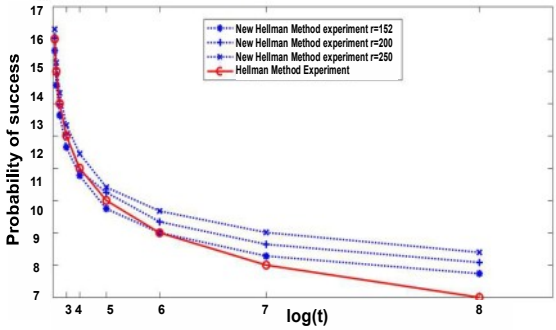
با توجه به نتایج جدول (۱) مشاهده می‌شود که اگر چه احتمال موفقیت در مقایسه با روش هلمن افزایش بسیار خوبی داشته است؛ اما نمی‌توانیم به‌طور هم‌زمان، هم زمان برخط و هم حافظه را کاهش دهیم. به‌عنوان نمونه، در سطر دوم جدول (۱) حافظه نسبت به روش هلمن تغییری نکرده است؛ اما زمان برخط افزایش یافته است. در سطر سوم نیز زمان برخط ثابت مانده است؛ اما حافظه افزایش یافته است.

مثال. برای مقایسهٔ تجربی احتمال موفقیت، حافظه و زمان برخط روش هلمن با روش ارائه‌شده در این بخش، مشابه با قبل، از الگوریتم mini AES استفاده می‌کنیم. ابتدا با استفاده از روش هلمن و شرط توقف ماتریس $mt^2 = N$ و همچنین شرط $t=r$، جداول هلمن بسیاری تولید می‌کنیم؛ سپس، میانگین احتمال موفقیت، حافظه و زمان برخط آنها را به‌دقّت بر حسب $\log(t)$ رسم کرده‌ایم. درنهایت، چنین نمودارهایی را نیز برای روش ارائه‌شده و با تعداد جداول متفاوت $r = 128, 200, 256$ رسم کرده‌ایم. نتایج در شکل‌های (۸، ۹، ۱۰) ارائه شده‌اند.



(شکل-۸): منحنی تجربی احتمال موفقیت $r=t$ ماتریس هلمن با شرط توقف ماتریس $N = mt^2$ (منحنی با علامت دایره) و منحنی‌های احتمال موفقیت روش ارائه‌شده به‌ازای $r = 150$ (منحنی نقطه‌چین با علامت ستاره)، $r = 200$ (منحنی نقطه‌چین با علامت مربع) و $r = 250$ (منحنی نقطه‌چین با علامت ضربدر) بر حسب $\log(t)$.

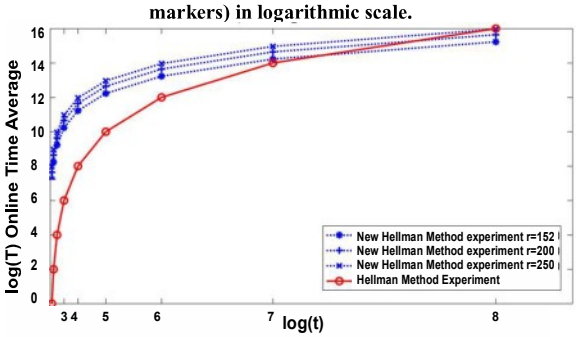
(Figure-8): Experimental probability of success of $r = t$ Hellman matrix with matrix stopping rule condition $N = mt^2$ (curve with circle markers); Experimental probability of success of r table with presented method for $r = 150$ (dotted curve with circle markers), $r = 200$ (dotted curve with square markers) and $r = 250$ (dotted curve with cross markers) in logarithmic scale.



(شکل-۹): منحنی تجربی واحد حافظه مورد نیاز برای ذخیره $r = t$ ماتریس هلمن با شرط توقف ماتریس $N = mt^2$ (منحنی

با علامت دایره) و منحنی تجربی واحد حافظه مورد نیاز برای ذخیره $r = 150$ (منحنی نقطه‌چین با علامت ستاره)، $r = 200$ (منحنی نقطه‌چین با علامت مربع) و $r = 250$ (منحنی نقطه‌چین با علامت ضربدر) بر حسب $\log(t)$ با روش ارائه‌شده.

(Figure-9): Experimental memory requirements of $r = t$ Hellman matrix with matrix stopping rule condition $N = mt^2$ (curve with circle markers); Experimental memory requirements of r table with presented method for $r = 150$ (dotted curve with circle markers), $r = 200$ (dotted curve with square markers) and $r = 250$ (dotted curve with cross markers) in logarithmic scale.



(شکل-۱۰): منحنی تجربی زمان برخط $r = t$ ماتریس هلمن با شرط توقف ماتریس $N = mt^2$ (منحنی با علامت دایره) و منحنی تجربی زمان برخط $r = 150$ (منحنی نقطه‌چین با علامت ستاره)، $r = 200$ (منحنی نقطه‌چین با علامت مربع) و $r = 250$ (منحنی نقطه‌چین با علامت ضربدر) بر حسب $\log(t)$ با روش ارائه‌شده.

(Figure-10): Experimental online time of $r = t$ Hellman matrix with matrix stopping rule condition $N = mt^2$ (curve with circle markers); Experimental online time of r table with presented method for $r = 150$ (dotted curve with circle markers), $r = 200$ (dotted curve with square markers) and $r = 250$ (dotted curve with cross markers) in logarithmic scale.

همان‌طورکه در این شکل‌ها مشاهده می‌کنید، برای $r = 200$ و $r = 250$، احتمال موفقیت بهتر است؛ اما برای $r = 150$ احتمال موفقیت کمتر از احتمال موفقیت روش هلمن است. از طرفی، دیگر در حالاتی که احتمال موفقیت بهبود یافته، یا تعداد واحد حافظه بیشتر شده است و یا زمان برخط. بنابراین نتیجه می‌گیریم، با ثابت‌نگه‌داشتن زمان برخط، اگرچه در روش ارائه‌شده احتمال موفقیت بیشتری نسبت به روش هلمن داریم و در مرحلهٔ برخط نیز با هشدار خطای بسیار کمتری مواجه خواهیم بود، اما مجبور به استفاده از واحد حافظه بیشتری نسبت به روش هلمن هستیم.

۶- نتیجه‌گیری

مصالحهٔ حافظه‌زمان یک الگوریتم احتمالاتی برای وارون‌کردن توابع یک‌طرفه است. هلمن یک کران پایین برای الگوریتم اصلی داده‌شده توسط خود ارائه داده است. تحلیل‌های سایر پژوهش‌گران برای اندازه‌گیری احتمال موفقیت نیز بر اساس همین کران پایین بوده است. در این مقاله، ابتدا نشان دادیم که یک شکاف بین کران پایین داده‌شده توسط هلمن و مقادیر تجربیِ به‌دست‌آمده توسط الگوریتم رمز mini AES وجود دارد. این شکاف بین نمودار تجربی و نظری در جایی که هر ماتریس هلمن تنها شامل یک زنجیره است، بیشتر از جاهای دیگر بود؛ لذا به بررسی ماتریس‌های هلمنی پرداختیم که دارای یک سطر هستند. نشان دادیم که امید ریاضی نرخ پوشش این‌گونه ماتریس‌های هلمن (با شرط توقف ماتریس) برابر با ۰/۸۵ است. در ادامه امید ریاضی نرخ پوشش برای یک ماتریس هلمن را به‌طور دقیق‌تری محاسبه کردیم. نتایج به‌دست‌آمده با مثال‌های عملی مقایسه شدند و نشان داده شد که فرمول پیشنهادی ما برای امید ریاضی نرخ پوشش، بسیار دقیق‌تر از کران پایین داده‌شده توسط هلمن است. درنهایت، ازآنجایی‌که احتمال موفقیت در مصالحهٔ حافظه‌زمان دارای اهمیت بسیاری در کارهای عملی است، روشی برای ساختن ماتریس‌هایی با سطح پوشش بالا ارائه کردیم. در این روش، ماتریس‌ها تنها از یک زنجیره با طول غیرثابت تشکیل شده‌اند. درحقیقت هر زنجیره تا قبل از رسیدن به نخستین تکرار ادامه می‌یابد. یکی از ویژگی‌های مهم این‌گونه ماتریس‌ها این است که در مرحلهٔ برخط هیچ‌گونه هشدار خطایی نداریم. در این بخش سعی کردیم علاوه‌بر این‌که ویژگی‌های خوب این‌گونه ماتریس‌ها (از جمله احتمال موفقیت و کاهش هشدار خطا) را حفظ کنیم، مشکل موجود در زمان برخط را نیز، با شکستن زنجیره‌ها به زیرزنجیره‌ها، تا حدّی جبران کنیم؛ اما نتایج نشان دادند که با این کار واحد حافظهٔ مورد نظر افزایش می‌یابد.

۷- مراجع 7- References

[1] E. Barkan, E. Biham and A. Shamir, “Rigorous bounds on cryptanalytic time/memory tradeoffs”, in *Advances in Cryptology: Proceedings of Crypto, LNCS*, 2006, 4117, pp. 1–21.

[2] J. Borst, “Block Ciphers: Design, Analysis and Side-Channel Analysis”, PhD thesis, Katholieke Universiteit Leuven, 2001.

[3] D.E. Denning, *Cryptography and data security*, Addison-Wesley, 1982.

[4] P. Flajolet, and A. Odlyzko, “Random mapping statistics”, *Advances in Cryptology, Proceedings of Eurocrypt’89, LNCS*, 1990, 434, pp. 329–354.

[5] P. Flajolet, and R. Sedgewick, *An introduction to analysis of algorithms*, Addison-Wesley, 2013.

[6] B. Harris, "Probability distributions related to random mappings", *The Annals of Mathematical Statistics,* pp. 1045–1062, 1960.

[7] M. Hellman, "A cryptanalytic time-memory trade-off", *IEEE Transactions on Information Theory IT*, vol. 26, pp. 401–406, 1980.

[8] J. Hong, and S. Moon, "A comparison of cryptanalytic tradeoff algorithms", *Journal of cryptology*, vol. 26 (4), pp. 559-637, 2013.

[9] J. Hong and B.I. Kim, "Performance comparison of cryptanalytic time memory data tradeoff methods", *Bull. Korean Math. Soc.*, vol. 53(5), pp. 1439-1446, 2016.

[10] V. Kolchin, "*Random mapp-ings*", Translations series in mathematics and engineering, Optimization Software, Inc., Publications Division, 1986.

[11] K. Kusuda, and T. Matsumoto, "Optimization of time-memory trade-off cryptanalysis and is application to DES", *FEAL-32, and Skipjack*, E-79A, pp. 35–48, 1996.

[12] G. W. Lee, and J. Hong, "Comparison of perfect table cryptanalytic tradeoff algorithms", *Designs, Codes and Cryptography*, pp. 473-523, 2016.

[13] R.C. Phan, "Mini advanced encryption standard (mini-AES): a testbed for cryptanalysis students", *Cryptologia,* vol. 26(4), pp. 283-306, 2002.

[14] Ph. Oechslin, "Making a faster cryptanaly-tic time-memory trade-off", *Annual International Cryptology Conference*, Springer Berlin Heidelberg, 2003.

ناصرحسین غروی، تحصیلات دانشگاهی خود در مقطع کارشناسی در رشتهٔ مهندسی مخابرات در سال ۱۳۷۵ از دانشگاه تهران، کارشناسی ارشد در رشتهٔ مخابرات رمز در سال ۱۳۸۰ و مدرک دکترای خود را در رشتهٔ مهندسی دفاع سایبری در سال ۱۳۹۶ از دانشگاه جامع امام حسین (ع) اخذ کرد. زمینه‌های پژوهشی مورد علاقهٔ ایشان، طراحی، ارزیابی، تحلیل و شکست الگوریتم‌ها و پروتکل‌های رمزنگاری و همچنین امنیت شبکه است.

نشانی رایانامهٔ ایشان عبارت است از:

hgharavi@ihu.ac.ir

عبدالرسول میرقدری، در سال ۱۳۶۵ مدرک کارشناسی در رشتهٔ آمار ریاضی و در سال ۱۳۶۸ مدرک کارشناسی ارشد خود را در رشتهٔ آمار نظری از دانشگاه شیراز و همچنین مدرک دکترای خود را در رشتهٔ آمار ریاضی در سال ۱۳۸۰ از دانشگاه شیراز دریافت کرد. وی اکنون عضو هیات علمی و دانشیار دانشکدهٔ مهندسی فناوری اطلاعات و ارتباطات دانشگاه جامع امام حسین(ع) است. زمینه‌های پژوهشی مورد علاقهٔ ایشان فرآیندهای تصادفی، نهان‌نگاری، رمزنگاری و تحلیل الگوریتم‌های رمز است.

نشانی رایانامهٔ ایشان عبارت است از:

amrghdri@ihu.ac.ir

محمد عبداللهی ازگمی، تحصیلات دانشگاهی خود را در مقاطع کارشناسی، کارشناسی ارشد و دکترا در رشته مهندسی کامپیوتر-نرم‌افزار به‌ترتیب در سال‌های ۱۳۷۱، ۱۳۷۵ و ۱۳۸۴ در دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف به پایان رسانده است. زمینه‌های پژوهشی مورد علاقهٔ ایشان، مدل‌سازی و ارزیابی سامانه‌های رایانه‌ای و امنیت شبکه است. از ایشان تاکنون مقالات متعددی در مجلات و همایش‌ها به چاپ رسیده است. دکتر عبداللهی ازگمی هم‌اکنون دانشیار گروه نرم‌افزار در دانشکدهٔ مهندسی کامپیوتر، دانشگاه علم و صنعت ایران است.

نشانی رایانامهٔ ایشان عبارت است از:

azgomi@iust.ac.ir

سید احمد موسوی، در سال ۱۳۸۵ مدرک کارشناسی خود را در رشته ریاضی محض از دانشگاه کاشان دریافت کرد. همچنین در سال ۱۳۸۸ مدرک کارشناسی ارشد و در سال ۱۳۹۳ مدرک دکترای خود را در رشته ریاضی از دانشگاه شهید باهنر کرمان اخذ کرد. زمینه‌های پژوهشی مورد علاقهٔ ایشان نظریه اطلاعات کوانتومی، نظریهٔ ماتریس‌ها، رمزنگاری و تحلیل الگوریتم‌های رمز است.

نشانی رایانامهٔ ایشان عبارت است از:

s.a.mousavi@math.uk.ac.ir