بررسی پیچیدگی خطی دنباله های تصادفی و بیان یک آزمون آماری

محمد دخیل علیان، محمدرضا عارف و محمود مدرس هاشمی

دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

تلفن ۸۹۱۲۴۵۸، نمابر ۸۹۱۲۴۵۱

پست الکترونیکی md-alian@cc.iut.ac.ir

چکیده: داشتن پیچیدگی خطی بزرگ یکی از ویژگیهای مطلوب دنباله های شبه تصادفی است. این شرط به تنهایی کافی نیست و علاوه برآن باید نمودار پیچیدگی خطی به صورت پله‌ای باشد. در این مقاله ضمن بیان ویژگی دنباله‌های کاملا تصادفی به منظور برآورده کردن یک روش عملی برای تطبیق یا عدم تطبیق توزیع پیچیدگی خطی دنباله‌های تصادفی با دنباله‌های شبه تصادفی، آزمون جدیدی پیشنهاد شده است.

کلمات کلیدی: دنباله‌های تصادفی، رمزکننده‌های پی‌درپی، پیچیدگی خطی، آزمون آماری

۱- مقدمه

در سیستمهای رمز "پی در پی" دنبالهٔ کلید اجرایی باید دارای خواص مطلوبی باشد. گالوب معیار داشتن دورهٔ تناوب بزرگ و شبه تصادفی بودن کلید اجرایی را طرح کرده است [۲]. واضح است که از میان کلیه دنباله‌ها، تعداد محدودی دارای خواص آماری خوب بوده و معیارهای گالوب را برآورده می‌سازند. در حال این معیارها، معیارهای لازمی هستند به این معنا که ممکن است مولد دنباله‌هایی، معیارهای گالوب را برآورده سازند ولی قابل استفاده نباشند. LFSR [1] نمونه‌ای از این مولدها است که بخوبی معیارهای گالوب را برآورده می‌سازد. با کمک این مولد می‌توان به دنباله‌هایی با دورهٔ تناوب بسیار بزرگ با خواص آماری مناسب دست یافت، با این وجود براحتی اثبات می‌شود که با داشتن 2n بیت از خروجی یک LFSR می‌توان حالت اولیه و چند جمله‌ای مشخصهٔ LFSR را بدست آورد. مشکل اصلی LFSR ها ناشی از خطی بودن آنهاست -که باعث می‌شود خروجی آنها کاملاً" قابل پیش‌بینی باشند. معیاری که با توجه به این مشکل طرح شده است: بزرگ بودن پیچیدگی خطی دنبالهٔ کلید اجرایی است. بدین جهت در بسیاری از مولدها با بکار گیری ساختارهای غیرخطی سعی می‌نمایند، پیچیدگی خطی را تا حد ممکن افزایش دهند ([۴];[۵]).

بزرگ بودن پیچیدگی خطی نیز به تنهایی کفایت نمی‌کند زیرا این شرط موقعی مطلوب است که پیچیدگی خطی به صورت پله‌ای باشد. در مقالهٔ حاضر به بررسی احتمالی پیچیدگی خطی دنباله‌های تصادفی خواهیم پرداخت. سپس با توجه به اینکه بخشهای مختلف کلید اجرایی نیز باید خواص مطلوبی ازاین دیدگاه داشته باشند، به منظور بررسی تطابق دنباله‌های شبه تصادفی با دنباله‌های کاملاً" تصادفی ضمن بیان یک قضیه، آزمون آماری جدیدی پیشنهاد می‌شود.

۲- پیچیدگی خطی دنباله‌های تصادفی

دنبالهٔ $S = 0,0,\ldots,0,1$ به طول n بیت را در نظر بگیرید. پیچیدگی خطی این دنباله برابر n بوده اما این دنباله به عنوان کلید اجرایی نمی‌تواند مورد استفاده قرار گیرد؛ زیرا بوضوح روشن است که خواص آماری مطلوبی ندارد. از این مثال در می‌یابیم که بالا بودن پیچیدگی خطی یک شرط لازم است و باید شرایط دیگری را نیز برای دنباله در نظر گرفت. شرط مهم دیگری که باید علاوه بر بزرگ بودن پیچیدگی خطی در نظر گرفت، پله‌ای بودن پیچیدگی خطی است. نمودار پیچیدگی خطی یک دنباله بداین صورت بدست می‌آید که از ابتدای دنباله

1- Linear Feedback Shift Register.

شروع کرده و پیچیدگی خطی مربوط به بیت اول، دو بیت اول، سه بیت اول و ... را پیدا می‌کنیم تا در نهایت پیچیدگی خطی کل دنباله بدست آید. پیچیدگی خطی را می‌توان توسط آلگوریتم برلکمپ-مسی [٦] بدست آورد. با بدست آمدن پیچیدگی خطی می‌توان نمودار پیچیدگی خطی را مقابل تعداد بیتهای متناظرش ترسیم نمود. این نمودار را LCP [2] می‌نامند. شرط پله‌ای بودن پیچیدگی خطی بدین معنی است که LCP مربوطه باید بصورت پله‌ای افزایش یابد. بنابراین دنبالهٔ S از این جهت که پیچیدگی خطی آن بصورت جهشی (و نه پله‌ای) به حداکثر خود می‌رسد دارای ضعف است.

برای اثبات لزوم پله‌ای بودن نمودار LCP در دنباله‌های کاملاً تصادفی دو قضیهٔ زیر مطرح است:

قضیه: میانگین وواریانس پیچیدگی خطی دنبالهٔ $S=s_0, s_1, \ldots, s_{n-1}$، که شامل n متغیر باینری تصادفی مستقل و یکنواخت (i.i.d) است عبارتند از [٤]:

$$E[\Lambda(S)]=\frac{n}{2}+\frac{4+R_2(n)}{18}-2^{-n}\left(\frac{n}{3}-\frac{2}{9}\right) \qquad (۱)$$

$$Var[\Lambda(S)]=\frac{86}{81}-2^{-n}\left(\frac{14-R_2(n)}{27}n+\frac{82+2R_2(n)}{81}\right)-2^{-2n}\left(\frac{n^2}{9}+\frac{4n}{27}+\frac{4}{81}\right) \qquad (۲)$$

($R_2(n)$ باقیماندهٔ تقسیم n بر ۲ می‌باشد.)

اگر n به مقدار کافی بزرگ باشد ($n>۱۰$)، میانگین و واریانس پیچیدگی خطی فوق برابر $\frac{n}{2}$ و $\frac{86}{81}$ خواهد شد. پس برای یک دنبالهٔ تصادفی به طول n انتظار این است که پیچیدگی خطی آن بسیار به $\frac{n}{2}$ نزدیک باشد. قضیهٔ دیگری که در اینجا مطرح می‌کنیم متوسط طول و ارتفاع پله‌ها را در نمودار LCP یک دنبالهٔ کاملاً تصادفی مشخص می‌کند.

قضیه: اگر $S=s_0, s_1, s_2, \ldots$ دنباله‌ای از متغیرهای تصادفی i.i.d باشد بطوریکه پیچیدگی خطی n بیت اول آن برابر با L باشد؛ در این صورت برای ایجاد یک پله در LCP، تعداد متوسط بیتهایی که باید به دنباله اضافه شود عبارت است از [٤]:

- میانگین طول پله

$$\begin{cases} 2 & \text{if} \quad L\le\frac{n}{2} \\ 2+2L-n & \text{if} \quad L>\frac{n}{2} \end{cases} \qquad (۳)$$

همچنین متوسط ارتفاع پله‌ها برابر است با:

- میانگین ارتفاع پله

$$\begin{cases} 2 & \text{if} \quad L\ge\frac{n}{2} \\ n-2L+2 & \text{if} \quad L<\frac{n}{2} \end{cases} \qquad (٤)$$

قضیهٔ دوم این حقیقت را روشن می‌سازد که مقدار پرش در LCP دنباله‌های تصادفی نامنظم می‌باشد. دنباله‌هایی نیز وجود دارند که دارای پیچیدگی خطی پله‌ای منظم هستند. این دنباله‌ها به صورت زیر تعریف می‌شوند:

تعریف: دنبالهٔ $S=s_0, s_1, s_2, \ldots$ دارای نمودار LCP کامل یا ایده‌آل است هرگاه:

$$\forall n\ge 1 \quad \Lambda(S)=\left[\frac{n+1}{2}\right] \qquad (٥)$$

که در آن Λ و [ ] مشخص کنندهٔ پیچیدگی خطی و جزء صحیح می‌باشد.

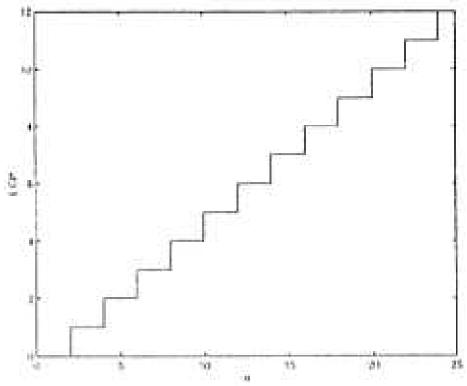
شکل (۱)، LCP یک دنبالهٔ با پیچیدگی خطی ایده‌آل را نشان می‌دهد. در چنین دنباله‌هایی معمولاً وابستگی زیادی بین بیتها وجود دارد و نمی‌توان از آنها در رمزنگاری استفاده کرد. به همین دلیل نمودار LCP دنباله‌های مورد استفاده نباید کاملاً منظم باشند، بلکه لازم است نامنظم بوده بطوری که میانگین پله‌ها و میانگین افزایش بیتها برای ایجاد یک پله با حالت ایده‌آل مطابقت داشته باشد. شکل (۲) نمونه‌ای از LCP یک دنبالهٔ تصادفی را نشان می‌دهد.

۳-معرفی یک آزمون آماری

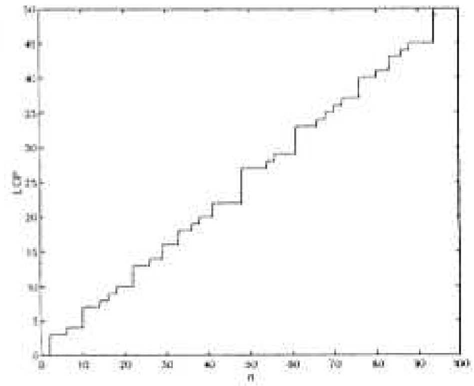
فرض کنید دنبالهٔ $S=s_0, s_1, s_2, \ldots$ یک دنبالهٔ باینری متشکل از n متغیر تصادفی i.i.d باشد. برای n های بزرگ ($n>۱۰$) مطابق رابطهٔ (۱)، میانگین پیچیدگی خطی برابر $\frac{n}{2}$ می‌باشد. از طرف دیگر با توجه قضایای

2-Linear Complexity Profile.

بخش قبل برای یک دنبالهٔ تصادفی، LCP باید بصورت پله‌ای باشد. اگر ..., h_2, h_1, $H=h_0$ دنبالهٔ مقدار پرش متناظر با هر بیت در LCP دنبالهٔ S باشد؛ در این حالت



شکل(۱): LCP یک دنبالهٔ ایده‌آل



شکل(۲): LCP دنبالهٔ نوعی کاملاً تصادفی

احتمال زیر (در حد) برای هر یک از پرشها در دنبالهٔ فوق قابل اثبات است[۷]:

$$P(h=m/h\neq 0)=\frac{1}{2^m} \qquad (۶)$$

با توجه به رابطهٔ (۶) می‌توان مقدار متوسط ارتفاع پله‌های غیر صفر را نیز در LCP بدست آورد:

$$E(h/h\neq 0)=\sum_{m=1}^{\infty} m\times P(h=m/h\neq 0)=\sum_{m=1}^{\infty}\frac{m}{2^m}=2$$

(۷)

از آنجا که بطور متوسط باید مجموع ارتفاع پله‌ها و متوسط پله‌های با ارتفاع غیر صفر برابر $\frac{n}{2}$ و ۲ باشد لذا باید بطورمتوسط تعداد پله‌های با ارتفاع غیر صفر برابر $\frac{n}{4}$ و تعداد پله‌های با ارتفاع صفر برابر $\frac{3n}{4}$ باشد. به عبارت دیگر می‌توان نوشت:

قضیه: اگر ..., s_2, s_1, $S=s_0$ یک دنبالهٔ تصادفی باینری با توزیع i.i.d باشد آنگاه (در حد) توزیع ارتفاع پرشها در نمودار LCP آن عبارت است از:

$$P(h=m)=\begin{cases}\frac{3}{4} & if \quad m=0\\ \frac{1}{2^{m+2}} & if \quad m=1,2,\ldots\end{cases}$$

(۸)

اثبات: دنبالهٔ $S_{m+n}=s_0, s_1, \ldots, s_{n-1}, s_n, \ldots, s_{m+n-1}$ و دنبالهٔ $H_{m+n}=h_0, h_1, \ldots, h_{n-1}, h_n, \ldots, h_{m+n-1}$ (دنبالهٔ مقدار پرش متناظر با هر بیت در LCP دنبالهٔ S) را در نظر بگیرید. طبق رابطهٔ (۱) برای n های به مقدار کافی بزرگ می‌توان نوشت:

$$\begin{aligned}E(\Lambda(S_{n+m})) &= E(h_0+h_1+\ldots+h_{n-1}+h_n+\ldots+h_{n+m})\\ &= E(\Lambda(S_n))+E(h_n+h_{n+1}+\ldots+h_{n+m})\\ &= \frac{n}{2}+\frac{4+R_2(n)}{18}+\varepsilon_n+m\times E(h)\end{aligned}$$

با بکارگیری تساوی $E(h)=E(h/h\neq 0)P(h\neq 0)$ و رابطه‌های (۱) و (۷) از تساوی فوق می‌توان $P(h\neq 0)$ را بدست آورد:

$$\frac{n+m}{2}+\frac{4+R_2(n+m)}{18}+\varepsilon_{n+m}=\frac{n}{2}+\frac{4+R_2(n)}{18}+\varepsilon_n+mE(h/h\neq 0)P(h\neq 0)$$

وبنا براین داریم:

$$P(h\neq 0)=\frac{1}{4}+\frac{R_2(m+n)-R_2(n)+\varepsilon}{36m}$$

(ε ، ε_n ، ε_{m+n} مقادیر کوچکی هستند که با بزرگ شدن m و n به سمت صفر میل می‌کنند)

برای m های بزرگ بخش دوم طرف راست عبارت فوق ناچیز شده و بنابراین داریم:

$$P(h=0)=1-P(h\neq 0)=1-\frac{1}{4}=\frac{3}{4} \qquad (۹)$$

از طرف دیگر با استفاده از رابطهٔ (۶) و (۹) می‌توان نوشت

$$\underset{m=1,2,3,\ldots}{P(h=m)}=P(h=m/h\neq 0)P(h\neq 0)+P(h=m/h=0)P(h=0) \quad :$$

$$=\frac{1}{2^m}\times\frac{1}{4}+0\times\frac{3}{4}=\frac{1}{2^{m+2}}$$

∎

با توجه به قضیهٔ فوق برای یک دنبالهٔ تصادفی به طول n بیت انتظار داریم بطور متوسط $\frac{3}{4}$ از h_i ها صفر، $\frac{1}{8}$ آنها یک، $\frac{1}{16}$ آنها دو و... باشد. برای انجام آزمون آماری پله‌ای بودن پیچیدگی خطی می‌توان از این نتایج استفاده نمود. به این ترتیب یک دنبالهٔ شبه تصادفی وقتی از جهت پله‌ای بودن پیچیدگی خطی مناسب است که آمارگان پله‌های موجود در LCP آن بسیار شبیه (۸) باشد.

۳-۱ آزمون χ^2

فرض کنید دنبالهٔ $S=s_0, s_1, \dots, s_{n-1}$ شامل n متغیر تصادفی باشد و هر کدام از s_i ها با توزیع نامشخصی یکی از مقادیر $\{a_0, a_1, \dots, a_m\}$ را اختیار کنند. فرض کنید احتمال P_j به صورت زیر تعریف شود:

$$P(S_i=a_j)=P_j \quad {}_{0\le i\le n-1,\ 0\le j\le m-1} \qquad (۱۰)$$

به عبارت دیگر همهٔ متغیرهای S_i دارای توزیع یکسانی باشند. در این صورت برای بررسی میزان تطبیق یا عدم تطبیق تابع توزیع فرض شده، مقدار χ^2 بصورت زیر تعریف می گردد[۸]:

$$\chi^2=\sum_{j=0}^{m-1}\frac{(N_j-nP_j)^2}{nP_j} \qquad (۱۱)$$

در عبارت فوق N_j تعداد دفعاتی است که a_j در دنبالهٔ S ظاهر شده است. هنگامی که n به سمت بینهایت میل کند، تابع توزیع χ^2 تقریبا" مستقل از P_j شده و در این حالت خواهیم داشت:

$$P(\chi^2\le k)=\int_0^k\frac{2^{-(\frac{m-1}{2})}\,y^{(\frac{m-3}{2})}\,e^{\frac{y}{2}}}{\Gamma(\frac{m-1}{2})}dy \qquad (۱۲)$$

($\Gamma(.)$ تابع توزیع گاما می‌باشد)

لازم به ذکر است که برای انجام آزمون باید رابطه (۱۳) برای تقریب هر چه بهتر توزیع کای-دو برقرار باشد.

$$\forall i \quad np_i \succ 5 \qquad (۱۳)$$

۳-۲ آزمون پله‌ای بودن پیچیدگی خطی

برای دنبالهٔ $S=s_0, s_1, s_2, \dots$ می‌توان با کمک آلگوریتم برلکمپ-مسی دنبالهٔ $H=h_0, h_1, h_2, \dots$ را بدست آورد. اگر S یک دنبالهٔ تصادفی با مؤلفه‌های i.i.d باشد در این صورت هر یک از h_i ها مقادیر ۰،۱،۲،... را با احتمالهای زیر تقریبا" اختیار می کنند:

$$P(h_i=m)=\begin{cases}\frac{3}{4} & if\ m=0\\ \frac{1}{2^{m+2}} & if\ m\neq 0\end{cases} \qquad i=0,1,2,\dots$$

(۱٤)

از آنجا که همهٔ متغیرهای h_i (در حالت حدی) دارای توزیع یکسان فوق هستند، در این صورت برای بررسی میزان تطبیق یا عدم تطبیق تابع توزیع (۱٤) با یک دنبالهٔ نمونه می‌توان از آزمون χ^2 مطابق با آنچه در بخش قبل مطرح گردید استفاده نمود.

فرض کنید دنبالهٔ نمونهٔ $S=s_0, s_1, \dots, s_{n-1}$ به طول n در اختیار است. ابتدا دنبالهٔ $H=h_0,h_1,h_2,\dots,h_{n-1}$ را بدست می‌آوریم و سپس با توجه به رابطهٔ (۱۱)، χ^2 را اینگونه تعریف می‌نماییم:

$$\chi^2=\frac{(Nh_0-\frac{3}{4}n)^2}{\frac{3}{4}n}+\sum_{i=1}^{m}\frac{(Nh_i-\frac{n}{2^{i+2}})^2}{\frac{n}{2^{i+2}}}$$

(۱۵)

در عبارت فوق Nh_i تعداد پله‌های به ارتفاع i در دنبالهٔ $H=h_0, h_1, h_2, \dots, h_{n-1}$ می‌باشد.

χ^2 محاسبه شدهٔ در(۱٤) دارای m درجهٔ آزادی است. با توجه به رابطهٔ (۱۳)، m باید به صورت زیر انتخاب نمود

$$m\prec\log_2(\frac{n}{20}) \qquad (۱٦)$$

مثال: اگر این آزمون را برای دنباله‌ای با رابطهٔ زیر-که LCP آن در شکل(۱) رسم شده است- اعمال کنیم:

$$S_n=\begin{cases}1 & if\ n=2^j-1 \quad j=0,1,2,\dots\\ 0 & otherwise\end{cases} \qquad (۱۷)$$

خواهیم داشت(با فرض اینکه n زوج باشد):

$$Nh_i = \begin{cases} \frac{n}{2} & if \quad i = 0,1 \\ 0 & otherwise \end{cases} \qquad (۱۸$$

با فرض $n=2^L$ و بکارگیری رابطه (۱۶) ، χ^2 برابر است با :

$$m = L-5$$

$$\chi^2 = 2^{L-2}\left(\frac{5}{6}+\sum_{i=2}^{L-5}\frac{1}{2^i}\right) \succ L-5 \qquad (۱۷)$$

اگر میزان اطمینان در آزمون فوق را ۹۵٪ در نظر بگیریم، ین دنباله هیچگاه از آزمون بیان شده عبور نخواهد کرد و این نتیجه بسیار مطلوب می باشد .

۴- خلاصه و نتیجه گیری

یکی از مشخصات دنباله‌های تصادفی داشتن LCP پله‌ای است. همانطور که ملاحظه نمودیم، نمودار پیچید گی خطی یک دنبالهٔ تصادفی باید بصورت پله‌ای باشد و علاوه بر آن باید ارتفاع و طول پله‌ها نامنظم وتصادفی باشد . در این مقاله سعی شده است با مروری بر خواص پیچید گی خطی دنباله‌های کاملا" تصادفی ، چگونگی LCP یک دنبالهٔ شبه‌تصادفی تبیین شود . در این راستا برای رسیدن به یک روش عملی برای انطباق یا عدم انطباق توزیع پیچید گی خطی یک دنباله نمونه با یک دنبالهٔ کاملا" تصادفی،با بیان قضیه‌ای ، یک آزمون آماری طرح گردید .این آزمون یک آزمون χ^2 می‌باشد که چگونگی استخراج آن در بخش ۳ توضیح داده شد . بکارگیری این آزمون بر روی دنباله‌های مختلف نتایج کاملا" رضایت بخشی را بدنبال داشته است[۱].

مراجع :

[۱] محمد دخیل‌علیان، " طراحی و ارزیابی دنباله‌های شبه تصادفی غیرخطی در سیستمهای رمزی دریی" ، رساله دکتری، دانشگاه صنعتی اصفهان، در حال تدوین.

[2] Golomb .S.W. ; Shift Register Sequences , Holden-Day , San Fransisco , 1982.

[3] Beker.H. Piper.F. ;Cipher System : The protection of Communication , London, 1982.

[4] Rueppel.R.A. ; Analysis and Design of Stream Cipher ; Springer- Verlag , 1986

[5] Golic ,"On The Linear Complexity of Function of periodic GF(q) Sequences ", IEEE Trans. on Inf. Theory , Vol. 35 , No.1 Jan.1989.

[6] Massey . J.L ," Shift Register Synthesis and BCH Decoding ", IEEE Trans. on Inf. Theory ,Vol. 15 , No.1,pp.122-127 Jan.1969.

[7] Niederriter.H ." The Probabilistic Theory of Linear Complexity ",Springer- Verlag Advances in Cryptology, Eurocryp' 88 pp.191-209, 1988.

[8] Knuth : The Art of Computer Programming Vol.2 Addison-Wesley ,1981.