اصلاح آزمون رنها وارائه آزمون رنهابرای زیر قالبها

بهروز پی زری(دانشجوی کارشناسی ارشدمخابرات دانشگاه صنعتی اصفهان)
b_peyzari@hotmail.com

محمد دخیل علیان(استادیار دانشگاه صنعتی اصفهان)
mdalian@cc.iut.ac.ir

چکیده:تولید و بررسی دنباله های شبه تصادفی یکی از مباحث مهم در علم رمزنگاری می باشد.جهت بررسی خواص تصادفی دنباله های شبه تصادفی از تئوری آزمونهای آماری استفاده می شود.در آزمونهای آماری هدف ارائه ملاک عملی برای سنجش خواص دنباله مورد آزمون ومیزان نزدیکی آن به دنباله ایده ال یعنی دنباله های با توزیع یکنواخت که اعضای آن تؤاما مستقل هستند،می باشد.در این مقاله با نگرشی دقیق به خاصیت رنها در یک دنباله مدل احتمال دقیق آن را بدست می آوریم و به ارائه آزمونی دقیق برای بررسی تعداد رنها در یک دنباله شبه تصادفی می پردازیم .در ادامه نیز با بررسی مدل رنها برای زیر قالبها ی یک دنباله آزمون رنها برای زیر قالبها را ارائه خواهیم داد.

واژگان کلیدی:توزیع رنها[1] ،دنباله های شبه تصادفی،آزمونهای آماری[2]

۱.مقدمه

در بررسی و تجزیه و تحلیل الگوریتمهای رمز قالبی و سیستمهای رمز پی در پی دو مساله از اهمیت خاصی بر خوردار هستند.یکی از این مسایل بررسی تحلیلی این گونه سیستمها مانند مقاومت در برابر حمله تفاضلی ،حمله خطی و... می باشد. دیدگاه دیگر که از اهمیت و جایگاه خاصی در رمزنگاری و دیگر شاخه های علوم برخور دار است بررسی آماری قالبها و دنباله های تولید شده در این گونه سیستمها می باشد.در آزمونهای آماری میزان نزدیکی این دنباله های شبه تصادفی با حالت ایده ال مورد بررسی قرار می گیرد. در آزمونهای آماری با بدست آوردن مدل احتمال برای ویژگیهایی چون تعداد یکها،تعداد رنها،تعداد بلوکهاو گپها،میزان پیچیدگی ،میزان فشرده شدن دنباله و... میزان نزدیکی دنباله تولیدشده را با حالت ایده ال مورد

[1] Runs
[2] Statistical Testing

سنجش قرار می دهندو در نهایت با بدست آوردن آماره آزمون[1] و مقدار احتمال[2] و تعیین سطح معنی داری[3] که توسط آزمون کننده مشخص می شود به تصمیم گیری در مورد دنباله در میزان نزدیکی آن به دنباله ایده ال در خاصیت مذکور مبادرت می شود. آزمون رنها یکی ازآزمونهای ارئه شده توسط NIST می باشد [۱] که آماره آزمون آن به صورت نادقیق می باشد.در این مقاله به طور دقیق به خاصیت رنها برای کل دنباله و زیر قالبهای آن پرداخته شده است وآزمونهای آماری جدیدی با آماره های نسبتا دقیقتر در این زمینه ارائه خواهد شد.

۲.دو قضیه اساسی

قضیه۲-۱: قضیه زیبندگی کای دو[۲] :اگر $(X_1, X_2, X_3, \ldots, X_n)$ نمونه های مستقل متغیر تصادفی چند جمله ای X با پارامتر های $n, p_1, p_2, \ldots, p_m$باشد که در آن X_i $(i = 1,2,\ldots,n)$ هر یک از مقادیر a_j $(j = 1,2,\ldots,m)$ را با احتمال زیر بگیرند:

$$p(X_i = a_j) = p_j \qquad ,j = 1,2,\ldots,m \qquad ,i = 1,2,..,n$$

در این صورت اگر n به اندازه کافی بزرگ باشد عبارت زیر به سمت توزیع کای دو با $m-1$ درجه آزادی میل خواهد کرد.

$$T_n(obs) = \sum_{i=1}^{m} \frac{(N_i - np_i)^2}{np_i} \qquad (۱)$$

که در آن منظور از N_i تعداد دفعاتی است که a_i در متغیر تصادفی چند جمله ای نمونه ظاهر شده است.نکته مهم دیگر در استفاده از این قضیه بر قرار بودن شرط زیر است تا تقریب توزیع کای دو مناسب باشد:

$$\forall i \qquad np_i \geq 5 \qquad ,i = 1,2\ldots,m$$

قضیه۲-۲: قضیه خود همبستگی دنباله های $i.i.d$ با توزیع یکنواخت[۴] :اگر $S^n : s_1, s_2, \ldots, s_n$ یک دنباله تصادفی باینری با مؤلفه های مستقل و با توزیع یکنواخت باشند،آنگاه دنباله C^τ که به عنوان دنباله خود همبستگی مرتبه τ ام دنباله نا میده می شودنیزدنباله ای با مولفه های مستقل و با توزیع یکنواخت می باشند.

---

[1] Test Statistic
[2] Probability Value
[3] Level of Signficance

$$C^{\tau} : s_1 \oplus s_{\tau+1}, s_2 \oplus s_{\tau+2}, \ldots, s_{n-\tau} \oplus s_n = c_1^{\tau}, c_2^{\tau}, \ldots, c_{n-\tau}^{\tau}$$
$$\tau = 1,2,\ldots,n-1$$

۳.بدست آوردن مدل رنها به صورت دقیق و ارائه آزمون آماری بر اساس آن

گذر از یک به صفر و از صفر به یک در دنباله را معادل وقوع یک رن می نامیم. در واقع در آزمون رنها هدف بررسی تعداد این نوسانها می باشد.آماره آزمون در [۱] به صورت تقریبی بدست آمده است. در این آزمون توزیع رنها به صورت گوسی با میانگین و واریانس زیر در نظر گرفته شده است:

$$\mu = 1 + \frac{2n_0 n_1}{n} \quad (۲)$$

$$\sigma^2 = \frac{(\mu-1)(\mu-2)}{n-1} \quad (۳)$$

آنچه در زیر می آید بررسی دقیقتر رنها و بدست آوردن آماره آزمونی دقیق می باشد .برای بدست آوردن آماره آزمون، دنباله زیر را در نظر بگیرید:

$$S^n : s_1, s_2, \ldots, s_n$$

دنباله خود همبستگی مر تبه اول دنباله مذکور به صورت زیر می باشد:

$$C^1 : s_1 \oplus s_2, s_2 \oplus s_3, \ldots, s_{n-1} \oplus s_n = c_1, c_2, \ldots, c_{n-1}$$

که دارای طول $n-1$ است.با توجه به قضیه (۲-۲)اگر S^n یک دنباله با توزیع یکنواخت و تؤا ما مستقل باشد، C^1نیز همین خاصیت را داراست.حال نکته اساسی در بدست آوردن آماره آزمون رنها این مطلب است که هر بیت یک در C^1 معادل وقوع یک رن در S^n یعنی دنباله اصلی می باشد.بنابراین می توان گفت که عبارت زیر نمایانگر تعداد کل رنها در دنباله می باشد:

$$R = \sum_{i=1}^{n-1} c_i \quad (۴)$$

در بعضی از مراجع مثل [۱] مقدار $R+1$ را به عنوان تعداد کل رنها تعریف می کنند.ما در ادامه از این تعریف استفاده خواهیم کرد هر چند که در روند ارائه آزمون تفاوتی ایجاد نمی کند.حال چون c_i ها تؤاما مستقل با توزیع یکنواخت می باشد میانگین و واریانس تعداد رنها $(R+1)$ به صورت دقیق زیر بدست می آید:

$$E(R+1) = \sum_{i=1}^{n-1} E(c_i) + 1 = (\sum_{i=1}^{n-1} \frac{1}{2}) + 1 = \frac{(n-1)}{2} + 1 = \frac{n+1}{2} \quad (۵)$$

$$VAR(R+1) = VAR(R) = VAR(\sum_{i=1}^{n-1} c_i) = \sum_{i=1}^{n-1} VAR(c_i) = \sum_{i=1}^{n-1} (\frac{1}{2} - \frac{1}{4}) = \frac{n-1}{4} \quad (۶)$$

حال با توجه به قضیه حد مرکزی توزیع $R+1$ درn های به اندازه کافی بزرگ دارای توزیع نرمال با میانگین و واریانس بالا خواهد بودو بنابراین آماره آزمون که به صورت دقیق زیر بدست می آید دارای توزیع نرمال استاندارد خواهد بود:

$$T_n(obs) = \frac{R_{obs} - \frac{(n+1)}{2}}{\sqrt{\frac{n-1}{4}}} \quad (۷)$$

که در آن R_{obs} تعداد رنهای مشاهده شده در دنباله مورد آزمون می باشد که طبق تعریف پذیرفته شده بدست می آید.در نهایت با توجه به تئوری آزمون فرض مقدار احتمال به صورت زیر قابل محاسبه است:

$$P\_value = 2(1 - \Phi_n(|T_n(obs)|)) \quad (۸)$$

که در آن $\Phi_n(.)$ تابع توزیع نرمال استاندارد می باشد.در واقع P_value دو برابر مساحت زیر منحنی تابع چگالی نرمال استاندارد از $|T_n(obs)|$ تا بینهایت می باشد.[۱و۵]

برای انجام آزمون کافی است با توجه به رابطه بالا P_value محاسبه گردیده و با سطح معنی داری α $(0.001 \le \alpha \le 0.05)$مقایسه گرددودر نهایت از دو گزاره زیر برای تصمیم گیری در مورد آزمون استفاده شود:[۱،۵]

الف:اگرمقداراحتمال محاسبه شده بزرگتریا مساوی مقدارانتخابی α باشد، از دیدگاه این آزمون دنباله تصادفی تلقی می شود.

ب:اگر مقداراحتمال محاسبه شده کوچکتر از مقدار انتخابی α باشد، از دیدگاه این آزمون دنباله تصادفی تلقی نمی شود.

شرط آزمون بزرگ بودن n به اندازه کافی است که برای این منظور در عمل $n \ge 20$ کفایت می کند.با همین ایده می توان آزمون رنها را به زیر قالبها نیز تعمیم داد.

۴. مدل رنهابرای زیر قالبها و ارائه آزمونی آماری براساس آن

همانطور که گفته شد در آزمون رنها هدف بررسی تعداد کل نوسانهای از صفر به یک و از یک به صفر در دنباله می باشد.این ایده را می توان با بدست آوردن مدل رنها برای یک قالب به طور کلی به زیر

قالبها نیز تعمیم داد و در نهایت مانند قسمت قبل با بدست آوردن آماره آزمون و محاسبه P_value انجام آزمون را میسر ساخت.

لم: در یک قالب به طول L احتمال وقوع i رن برابر است با عبارت زیر:

$$P(R=i)=\frac{\binom{L-1}{i-1}}{2^{L-1}}, i=1,2,\ldots,L \qquad (۹)$$

اثبات: احتمال وقوع i رن در یک قالب به طول L برابر است با احتمال وقوع $i-1$ تا بیت یک در قالب مربوط به دنباله خود همبستگی مرتبه اول آن، یعنی برابر است با احتمال وقوع $i-1$ تا یک در یک قالب به طول $L-1$ ، که باتوجه به اینکه در حالت ایده ال وقوع یک قالب به طول $L-1$ با $i-1$ تا یک برابر با ترکیب $i-1$ از $L-1$ است و تعدادکل قالبها به طول $L-1$ برابر با 2^{L-1} است،احتمال مطلوب برابر با عبارت زیر خواهد بود:

$$p_i=\frac{\binom{L-1}{i-1}}{2^{L-1}}, \forall i, i=1,2,\ldots,L \qquad (۱۰)$$

■

حال برای انجام آزمون آماره آزمون را با توجه به قضیه(۲-۱) به صورت زیر تشکیل می دهیم که در آن $N=\frac{n}{L}$ تعداد قالبها می باشد:

$$T_n(obs)=\sum_{i=1}^{L}\frac{(R_i-N.\frac{\binom{L-1}{i-1}}{2^{L-1}})^2}{N.\frac{\binom{L-1}{i-1}}{2^{L-1}}} \qquad (۱۱)$$

که در آن R_i برابربا تعداد قالبهایی است که دارای i رن می باشند.در اینجا آماره آزمون دارای توزیع کای دو با $L-1$ در جه آزادی خواهد بودو بنابراین مقدار P_value به صورت زیر بدست خواهد آمد[۱و۵]:

$$P\_value=gamma(L-1,T_n(obs)) \qquad (۱۲)$$

درواقع P_value برابر با مقدار مساحت زیر منحنی تابع چگالی توزیع کای دو از $T_n(obs)$ تا بی نهایت می باشد.در نهایت با انتخاب α مانند قسمت قبل می توان به تصمیم گیری در مورد دنباله مورد آزمون پرداخت.

شرط آزمون با توجه به قضیه (۲-۱) برقرار بودن رابطه زیر می باشد:

$$N.\frac{\binom{L-1}{i-1}}{2^{L-1}} \geq 5, \forall i, i = 1,2,\ldots, L \qquad (۱۳)$$

نکته ای که در پایان این آزمون قابل ذکر است این مطلب می باشد که چون در آزمون رنها به نحوی خود همبستگی مرتبه اول دنباله مورد بررسی قرار می گیردو در آزمون رنها برای زیر قالبها خود همبستگی مرتبه اول زیر قالبها بررسی می شود،می توان با بدست آوردن دنباله خود همبستگی مراتب بالاتر و انجام آزمون رنها بر روی زیر قالبهای آن به نحوی خود همبستگی زیر قالبهای با مراتب بالاتر از یک را نیز مورد آزمون قرار داد.

۵.نتیجه گیری

در این مقاله با نگرشی دقیق به مسا له رنها مدل رنها را به صورت نسبتا دقیقتر بدست آوردیم و با توجه به آن آزمون آماری رنها را به صورت دقیق بیا ن نمو دیم و در ادامه با بدست آوردن مدل رنها برای زیر قالبها آزمون جدید ی در این زمینه برای سنجش تصا دفی بودن زیر قالبها در یک دنباله ارائه داده شد که می تواند در بررسی آماری سیستمهای رمز نگاری مخصوصا سیستمهای رمز نگاری قالبی از آن استفاده نمود.

مراجع

[1]Rukhin A.L and Soto J. and Nechvatal J.,...,"A Statistical Test Suite for Random and Pseudo Random Number Generators for Cryptographic Applications"*NIST Special Publication 800-22*, 15 may ,2001

[2]Rohatgi V.K.," *Introduction to Probability Theory and Mathematical Statistics*",Johnwiley and sons,1976

[3]Knuth D."*The Art of Computer Programming*",vol. II,Addison Weslley Publishing Company,second edition,1983

[۴]دخیل علیان محمد؛*ارزیابی دنباله های شبه تصادفی و طراحی مولد های آشوبی*؛رساله دکتری در رشته مهندسی برق دانشگاه صنعتی اصفهان،آبان ماه ۱۳۷۷

[5]Rukhin A.L,"Testing Randomness: A Suite of statistical Procedure",*Department of Mathematics and Statistics*,Baltimor, USA,2001

[6]Menzes Alfred," *Handbook of Applied Cryptography*",CRC Press ,1997