
بسم الله الرحمن الرحيم

دانشگاه رازی

دانشکده علوم پایه
گروه ریاضی

پایان‌نامه جهت اخذ درجه کارشناسی ارشد رشته ریاضی محض گرایش جبر

عنوان :

رمزنگاری تصویر رنگی مبتنی بر DNA و توابع آشوب

استاد راهنما:

دکتر مهرداد احمدزاده

نگارش:

ابراهیم صفری

مهر ۱۳۹۲

دانشگاه رازی

دانشکده علوم پایه
گروه ریاضی

پایان‌نامه جهت اخذ درجه کارشناسی ارشد رشته ریاضی محض گرایش جبر

نام دانشجو:

ابراهیم صفری

تحت عنوان :

رمزنگاری تصویر رنگی مبتنی بر DNA و توابع آشوب

در تاریخ توسط هیأت داوران زیر بررسی و با درجه به تصویب نهایی رسید.

۱- استاد راهنمای پایان‌نامه دکتر مهرداد احمدزاده‌راجی با مرتبه‌ی علمی استادیار امضاء:

۲- استاد داور داخل گروه دکتر بهروز عدالت‌زاده با مرتبه‌ی علمی استادیار امضاء:

۳- استاد داور خارج گروه دکتر محمود احمدی با مرتبه‌ی علمی استادیار امضاء:

سپاس گزاری...

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمت‌های او ندانند و کوشندگان، حق او را گزاردن نتوانند. و سلام و دورد بر محمّد و خاندان پاك او، طاهران معصوم، هم آنان که وجودمان وامدار وجودشان است، و نفرین پیوسته بر دشمنان ایشان تا روز رستاخیز. بدون شک جایگاه و منزلت معلم، اجّل از آن است که در مقام قدردانی از زحمات بی شائبه‌ی او، با زبان قاصر و دست ناتوان، چیزی بنگاریم. اما از آنجایی که تجلیل از معلم، سپاس از انسانی است که هدف و غایت آفرینش را تامین می‌کند و سلامت امانت‌هایی را که به دستش سپرده‌اند، تضمین، بر حسب وظیفه و از باب " من لم یشکر المنعم من المخلوقین لم یشکر اللّه عزّ و جلّ" : ازپدر و مادر عزیزم، این دو معلم بزرگوارم، که همواره بر کوتاهی و درشتی من، قلم عفو کشیده و کریمانه از کنار غفلت‌هایم گذشته‌اند و در تمام عرصه‌های زندگی یار و یاوری بی چشم داشت برای من بوده‌اند، از استاد با کمالات و شایسته، جناب آقای دکتر مهرداد احمدزاده راجی که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این رساله را بر عهده گرفتند، کمال تشکر و قدردانی را دارم. باشد که این خردترین، بخشی از زحمات آنان را سپاس گوید .

تقدیم به

پدر و مادر مهربانم و همسر عزیز و فداکارم که در سایه همیاری و همدلی او به این منظور نائل شدم، و تقدیم به تمام آزادمردانی که نیک می‌اندیشند و عقل و منطق را پیشه خود نموده و جز رضای الهی و پیشرفت و سعادت جامعه، هدفی ندارند. دانشمندان، بزرگان، و جوانمردانی که جان و مال خود را در حفظ و اعتلای این مرز و بوم فدا نموده و می‌نمایند.

چکیده

در این تحقیق سیستم رمزنگاری تصویر رنگی ارائه شده به وسیله Liu و همکاران و همچنین سیستم رمزنگاری Wei و همکاران مورد بررسی و تحلیل قرار گرفته است و اشکالات و نقایصی را که در پیاده‌سازی این دو الگوریتم وجود دارد و یا کارایی و امنیت آن‌ها را تهدید می‌کند، تصحیح کرده و در ادامه با پیاده‌سازی الگوریتم‌های اصلاح شده و انجام آزمون‌های استاندارد ارزیابی امنیت و کارایی، از امنیت و کارآمدی الگوریتم‌های اصلاح شده اطمینان می‌یابیم. در نهایت، الگوریتم رمزنگاری جدیدی با استفاده از توالی DNA و نگاشت لجستیک[۱] و سیستم لورنز[۲] پیشنهاد شده است که به خوبی آزمون‌های استاندارد ارزیابی امنیت و کارایی را می‌گذراند. علاوه بر امنیت مزیت دیگر الگوریتم پیشنهادی، سرعت اجرای آن نسبت به الگوریتم‌های اصلاح‌شده قبلی است. همچنین به علت استفاده از دو نگاشت آشوب می‌توان قسمتی ازاجرای الگوریتم را به صورت موازی انجام داد که همین باعث می‌شود زمان اجرای الگوریتم بسیار کاهش یابد.

واژه‌های کلیدی:رمزنگاری، نگاشت آشوب، جمع DNA، کدگذاری DNA

[۱] Logistic map
[۲] Lorenz system

فهرست مطالب

عنوان صفحه

[۱] Differential Attacks
[۲] Hamming distance

فهرست تصاویر

عنوان صفحه

فهرست جداول

عنوان صفحه

فصل ۱

نگرشی به آنچه خواهد آمد

۱.۱ مقدمه

امنیت یکی از ارکان موجودات زنده و احساس امنیت یکی از اساسی‌ترین نیازهای نوع بشر است. امروزه با گسترش وسایل ارتباطی و حجم اطلاعات مبادله شده در شبکه‌های رایانه‌ای و همچنین توسعه و پیشرفت‌های صنعت مخابرات چندرسانه‌ای، مفهوم مخابرات تصویری متحول شده است. امنیت رسانه‌های دیجیتال یکی از مسائل مهم و مطرح جامعه رمزنگاری در دنیای امروز است. با توجه به کاربرد روزافزون رایانه و گسترش زیرساخت‌های ارتباطی مثل شبکه‌های سیار و اینترنت، حفظ محرمانگی و تایید صحت تصاویر روز به روز اهمیت بیشتری می‌یابد. یک روش کارامد برای حفظ محرمانگی تصاویر، رمزنگاری آنها قبل از ارسال روی شبکه می‌باشد. به این ترتیب، فقط عوامل مجاز در صورت داشتن کلید صحیح قادر به رمزگشایی آن‌ها می‌باشند. تصویر خاکستری، یک داده حجیم دوبعدی است که کوچکترین واحد آن یک پیکسل است. هر پیکسل از یک تصویر دیجیتال، معرف میزان روشنایی آن نقطه از تصویر است. با توجه به میزان حساسیت چشم انسان در تشخیص سطوح روشنایی [۱] از یکدیگر، کل محدوده روشنایی قابل نمایش به ۲۵۶ سطح تقسیم‌بندی می‌شود. بنابراین سطح روشنایی هر پیکسل می‌تواند مقداری بین ۰ و ۲۵۵ داشته باشد. این محدوده توسط یک بایت قابل بازنمایی است. بنابراین یک تصویر خاکستری با ابعاد ۲۵۶×۲۵۶ پیکسل تقریباً معادل ۶۵ کیلوبایت است. پس یک تصویر با ابعاد کوچک، حجم اطلاعاتی بزرگی دارد، لذا با توجه به ویژگی‌های ذاتی داده‌های چندرسانه‌ای، وجود محدودیت در توان محاسباتی پردازنده، پهنای باند شبکه انتقال داده و زمان محاسباتی، باید از طرح‌های رمزنگاری کارآمد و متناسب با شرایط ذکرشده استفاده کرد. از دهه ۱۹۹۰ تا کنون، سامانه‌های دینامیک آشوبی به دلیل خصوصیاتی مثل حساسیت به شرایط اولیه و ارگادیک [۲] بودن نمو آن‌ها، به صورت گسترده در طراحی استراتژی‌های جدید برای رمزنگاری اطلاعات استفاده شده‌اند. در این تحقیق دو روش رمزنگاری تصویر را که در آن علاوه بر بکار بردن نگاشت‌های آشوب‌گون از توالی DNA استفاده می‌شود مورد بررسی و مطالعه قرار می‌دهیم. استفاده از توالی DNA به دلیل پیچیدگی‌های بیولوژیکی مسائل مربوط به آن و عملیات‌های جبری وابسته که بر اساس قوانین بیولوژیکی رشته‌های DNA طراحی و توسط پژوهشگران ارائه شده‌اند، پیچیدگی و امنیت الگوریتم رمزنگاری را افزایش می‌دهد. الگوریتم ارائه شده توسط Liu و همکاران [۱]، یک روش رمزنگاری تصویر رنگی است که در آن از جمع DNA و همچنین نگاشت آشوب لجستیک استفاده می‌شود که جزئیات آن در فصل دوم آمده است، اما پیاده‌سازی این الگوریتم و انجام آزمون‌های استاندارد امنیت و کارایی نشان می‌دهد که در این روش رمزنگاری میزان حساسیت به کلید پایین است و لذا در مقابل

[۱] Level of Intensity (Gray Scale)
[۲] Ergodic

حملات جستجوی جامع [1] مقاومت کمتری دارد. در حمله جستجوی جامع، مهاجم تلاش می‌کند تمام کلیدهای محتمل بر روی تصویر رمزشده را آزمایش کند تا بلکه متن یا تصویر اصلی بدست آید. بنابراین با ایجاد تغییرات و اصلاحاتی در الگوریتم، این نقص الگوریتم، که ضعیف بودن در مقابل حمله جستجوی جامع است را رفع کرده‌ایم.

روش دوم رمزنگاری که در این پایان‌نامه مورد مطالعه قرار گرفته، الگوریتم نوشته شده توسط Wei و همکاران [۲]، می‌باشد که مانند الگوریتم قبلی از کدگذاری DNA و عملیات جمع DNA در آن استفاده شده است اما سامانه آشوبی آن از معادلات سیستم ابر-آشوب [2] Chen بدست می‌آید. اما با مطالعه این الگوریتم رمزنگاری به این نتیجه می‌رسیم که این الگوریتم دارای نقایصی است که بدون اصلاح آن‌ها، پیاده‌سازی و اجرای الگوریتم غیرممکن است لذا با ایجاد اصلاحاتی در الگوریتم، پیاده‌سازی الگوریتم را عملی کرده‌ایم که جزئیات این روش و تغییرات ایجاد شده در فصل سوم آورده شده است. انجام آزمون‌های استاندارد از امنیت قابل قبول الگوریتم اصلاح شده حکایت دارد.

در ادامه یک روش رمزنگاری، طراحی و پیشنهاد شده است که برگرفته از مطالعه و بررسی دو روش رمزنگاری تصویر ذکر شده است که در آن از DNA و عملیات جمع DNA برای رمزنگاری و از تفریق DNA در رمزگشایی استفاده می‌شود. در این الگوریتم از نگاشت‌های آشوب لجستیک و معادلات آشوب لورنز برای تولید دنباله‌های آشوب استفاده شده است. آزمون‌های استاندارد، کارآمدی و امنیت بالای الگوریتم را نشان می‌دهد. نتایج اجرای الگوریتم‌ها در محیط متلب $(R۲۰۱۱a)$ ۷/۱۲/۰/۶۳۵، نشان از سرعت اجرای بالاتر الگوریتم پیشنهادی نسبت به دو روش قبلی دارد، همچنین استفاده از دو نگاشت آشوب بر پیچیدگی رمزنگاری می‌افزاید.

برای مطالعه این پایان‌نامه نیاز به مطالعه مقدمات و مفاهیمی از رمزنگاری و DNA و نیز نگاشت‌های آشوب‌گون است که در ادامه خواهد آمد.

۲.۱ مقدماتی بر رمزنگاری

کلمه Cryptography از یونانی قدیم گرفته شده است و آن ترکیب دو کلمه Krypto به معنی "مخفی" و کلمه grafo به معنی"نوشتن " است، بنابراین معنی تحت‌الفظی Cryptography "نوشته مخفی" است. رمزنگاری دانشی قدیمی از کدنویسی پیغام‌هاست که تنها فرستنده و گیرنده پیغام می توانند آن را درک کنند. رمزنگاری علمی است که از ریاضیات برای رمزی کردن و رمزگشایی اطلاعات استفاده می‌کند. رمزنگاری ما را قادر می‌سازد تا اطلاعات حساس را ذخیره و از میان یک شبکه ناامن مثل اینترنت انتقال دهیم بنابراین، کشف رمز به‌وسیله هر کس غیر از نامزد دریافت کننده پیام مشکل می‌باشد. رمزنگاری هنر نوشتن کدهای محرمانه و یک مهارت قدیمی است. اولین سندی که در آن از رمزنگاری استفاده شده به ۱۹۰۰ سال قبل از میلاد بر می‌گردد. بعضی از کارشناسان معتقدند که رمزنگاری ناگهان زمانی بعد از اختراع خط با بکاربردن نامه‌های تغییرشکل یافته رسمی دیپلماتیک در زمان جنگ و برای طرح نقشه جنگ پیدا شد. تعجب برانگیز نیست که شیوه‌های جدید رمزنگاری بعد از همگانی شدن ارتباطات کامپیوتری بوجود آمد. در انتقال داده‌ها و ارتباطات از راه دور،

[1] Brute-Force Attack
[2] hyper-chaotic

برای ارتباط روی هر رسانه ناامن که تقریباً شامل هر شبکه‌ای بویژه اینترنت است، رمزنگاری لازم وضروی است. در چارچوب یک ارتباط دو طرفه کاربردی، باید ابعاد امنیتی زیر لحاظ شود:

۱: احراز هویت [۱] : فرایند اثبات هویت یک شخص

۲: خصوصی و محرمانه ماندن اطلاعات [۲] : اطمینان از این‌که هیچ فردی غیر از دریافت‌کننده مورد نظر، نمی‌تواند پیام را بخواند.

۳: تضمین صحت اطلاعات [۳] : مطمئن شدن گیرنده پیام از اصل بودن پیام و اینکه به هیچ طریقی پیام تغییر نکرده است.

۴: غیر قابل انکار ساختن پیام‌ها [۴] : ساختاری که ثابت کند که واقعاً فرستنده، این پیام را فرستاده است.

۳.۱ انواع طرح‌های رمزنگاری

دو نوع طرح اصلی برای رمزنگاری وجود دارد:

۱: رمزنویسی متقارن: رمزنگاری متقارن یک نوع سیستم رمز است که در آن رمزنویسی و رمزگشایی، هر دو به‌وسیله یک کلید انجام می شوند. رمزنگاری متقارن، مستعد پذیرش حملات کشف رمز است و تجزیه و تحلیل کشف رمز آن خطی است یعنی قابل هک شدن هستند ورمزگشایی آن ساده است. این روش از نظر زمان لازم جهت کدگذاری و کدگشایی بسیار کارایی دارد زیرا زمان لازم برای این کار بسیار کم است. از سوی دیگر سطح محدودی از شناسایی طرف مقابل را فراهم می‌آورد البته این امر تا زمانی ممکن است که کلید بین طرفین محرمانه بماند زیرا تا زمانی‌که از محرمانه بودن این کلید مطمئن باشیم از اصالت طرف مقابل مطمئن هستیم ولی به مجرد لو رفتن کلید این امکان هست که شخص دیگری و با اطلاعات غلط نسبت به ارسال اطلاعات جعلی و از طرف شخص مورد نظر ما، به ارسال اطلاعات بپردازد. بنابراین قدرت سطح شناسایی طرف مقابل و اطمینان بخشی اصالت داده‌ها در آن بستگی به حفظ کلید دارد که این موضوع موجب ضعف این روش است.

۲: رمزنگاری نامتقارن: این روش رمزگذاری از مهمترین روشهای موجود می‌باشد که نقش اساسی در بسیاری از کاربردهای روزمره که دارای اهمیت بالایی هستند دارد بطوری‌که کاربرد آن در اطلاعات رمز و یا امضا شده بانکی و مالی از اهمیت بسیار بالایی برخوردار است. در مقایسه با روش کلید متقارن این روش دارای محاسبات بیشتر در زمینه کدکردن می‌باشد بنابراین زمان‌بر است و برای اطلاعات با حجم بالا مناسب نیست. کدگذاری به این شیوه با نام نامتقارن شناخته می‌شود زیرا بر خلاف روش متقارن در اینجا کلید کدگذاری و کدگشایی یکی نیستند زیرا در آن صورت به دلیل یکی بودن کلید در نزد دو طرف این

[۱] Authentication
[۲] Privacy/confidentiality
[۳] Integrity
[۴] Non-Repudiation

امکان بود که شخص ثالثی که کلید را کشف می‌کرد پیام‌ها را کدگشایی کند و یا حتی پیام‌های اشتباه برای طرفین ارسال کند و لذا طرفین رمزگذاری، به استناد این‌که فقط طرف مقابل آن‌ها از کلید مطلع است، قادر به تشخیص جعلی بودن پیام نخواهند بود. حال می‌توان به قوت این الگوریتم کدگذاری در تفکیک کلید کدگذاری و کلید کدگشایی پی برد بنابراین می‌توان گفت کدگذاری به شیوه کلید عمومی عبارت است از یک سیستم رمزگذاری که از دو کلید به نام عمومی که همه از آن مطلع هستند و کلید خصوصی که فقط گیرنده پیام از آن اطلاع دارد تشکیل شده است.

۴.۱ رمزنگاری DNA

رمزنگاری DNA یک حوزه جدید از رمزنگاری است در سالهای اخیر با پژوهش روی پتانسیل محاسباتی DNA به وجود آمده است. رمزنگاری DNA هنوز در گامهای نخستین خود می‌باشد و به همین دلیل تعداد مثال‌های اندکی در این حوزه ارائه شده است.
DNA به دلایل زیر، دارای این پتانسیل است که برای مدت مدیدی تبدیل به سیستم ذخیره اطلاعات متراکم شود.

۱: درعمل مقدار زیادی از داده‌ها می‌تواند با دقت بالایی در سطح مولکولی ذخیره شود. یک گرم DNA خشک (تقریباً به اندازه نصف يك حبه قند)، قابلیت ذخیره همان مقدار اطلاعات را دارد که می‌تواند روی یک تریلیون CD ذخیره کرد. تراکم مؤثر DNA حدود ۱۰۰۰۰۰ بار، بیشتر از هارد دیسک‌های مدرن است.

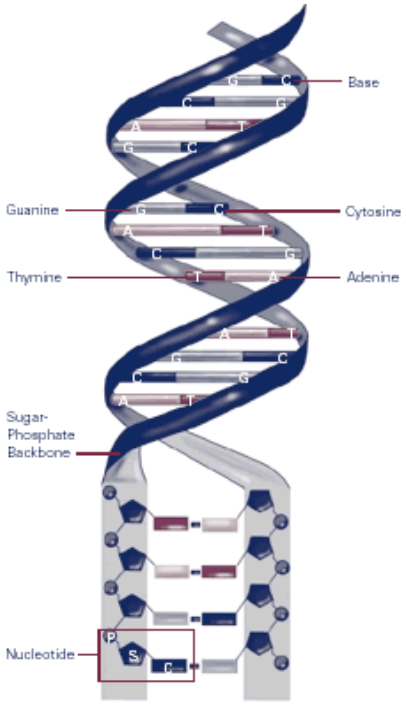
۲: با وجود تغییر شرایط محیطی و تکنولوژیکی، DNA همچون یک مخزن مولکولی برای میلیون‌ها سال باقی می‌ماند.

۳: اگر از DNA استفاده شود رونویسی کپی‌های کاملاً همانند از انباره اطلاعات، نسبتاً آسان است .

۱.۴.۱ ساختار DNA

در داخل سلول‌های هر موجود زنده ماده وراثتی بنام دئوکسی‌ریبونوکلئیک‌اسید [۱] (DNA) وجود دارد که به فرم یک مارپیچ دوگانه از نوکلئوتیدها است که حامل اطلاعات ژنتیکی از سلول هاست. تراکم داده‌ها در DNA شگفت انگیز است. DNA از ۳ جزءاصلی تشکیل شده است که عبارتند از قند، فسفات و باز آلی. هر رشته از DNA با استفاده از ۴ باز آلی به نام‌های آدنین (Adenine) و تیمین (Thymine) و سیتوزین (Cytosine) و گوانین (Guanine) کدنویسی شده است که این ۴ باز به طور مختصر با حروف A، T، C و G نشان داده می‌شوند. توالی این چهار باز آلی باعث رمزگذاری رشته ژنتیکی می‌شود که این رمزها برای ساخت اسید آمینه که واحدهای سازنده پروتئین می‌باشند مورد استفاده قرار می‌گیرد. یک مولکول DNA به عنوان وسیله ذخیره سازی اطلاعات ژنتیکی در هر بافت سلولی موجود زنده یافت می‌شود. در حالی که گروه‌های فسفات و قند نقش

[۱] Deoxyribonucleic Acid



شکل. ۱.۱ : ساختمان مولکول DNA

ساختمانی را دارا هستند. هر باز به یک مولکول قند و یک مولکول فسفات متصل است که با هم نوکلئوتید نامیده می‌شوند. طبق قانون مکملی واتسون- کریک A همواره با T پیوند دوگانه هیدروژنی دارد و C با G پیوند سه گانه دارد. نوکلوئوتیدها در دو رشته دراز به فرم مارپیچ مرتب شده‌اند که آن‌را مارپیچ دوگانه [۱] می‌نامیم. ساختار مارپیچ دوگانه تا حدودی مثل یک نردبان است که بازهای جفت شده پله‌های نردبان را شکل می‌دهند و مولکول‌های قند و فسفات پایه‌های قائم کناری را تشکیل می‌دهند. شکل ۱.۱، تصویری از یک مارپیچ دوگانه را نشان می‌دهد.

۲.۴.۱ ارتباط DNA با سیستم‌های کامپیوتری

آدلمن [۲] در سال ۱۹۹۴ ایده حیرت‌آور خود را مطرح کرد، او به این نتیجه رسید که مولکول‌های DNA دارای پتانسیل محاسباتی هستند، او توانست توانایی مولکول DNA را در حل مسائل پیچیده ریاضی به اثبات رساند.آدلمن به کمک تکنیکی جالب توانست برای یکی از مشهورترین مسائل محاسباتی، یعنی مسئله مسیر همیلتونی [۳] یا همان مسئله فروشنده دوره گرد [۴] راه حلی پیدا کند. موفقیت آدلمن توجهات بسیاری را در چند سال آخر قرن بیستم به خود جلب کرد. Gehani و همکاران [۳] در سال ۲۰۰۰ یک الگوریتم پنهان‌سازی

[۱] Double Helix
[۲] Adleman
[۳] HP /Hamiltonian Path Directed
[۴] TSP /Traveling Salesman Problem

تصویر، تحت رمزنگاری One-Time-Pad مبتنی بر DNA را طراحی کردند.

بعد از سال ۲۰۰۰ علاقه در رمزنگاری DNA محدود شد و یکی از دلایل آن این بود که دستکاری مولکول‌های DNA حتی برای بیوشیمیست‌ها و بیولوژیست‌ها خیلی مشکل بود، درسال‌های اخیر مجدداً دلبستگی‌هایی در این زمینه به وجود آمده است.

بر طبق مدل واتسون- کریک، میزان آدنین و تیمین برابر هستند زیرا بازهای آدنین در یکی از دو رشته، همیشه به تیمین رشته مقابل می‌پیوندد. به طور مشابه، میزان گوانین با سیتوزین نیز برابر است زیرا دو باز در مولکول DNA، همواره به هم پیوند می‌خورند. مدل واتسون - کریک نشان داد که اطلاعات ژنتیکی به نحوی در توالی بازهای مولکول DNA رمزشده است، همانند آن چه که در کامپیوترها اتفاق می افتد، یعنی ذخیره داده ها به صورت رشته‌های دودویی [۱] متشکل از دو رقم ۰ و ۱ میباشد. یک رقم دودویی، بیت [۲] خوانده میشود. اطلاعات در کامپیوترهای دیجیتال، به‌وسیله گروههایی از بیت نشان داده میشوند. با استفاده از تکنیک های کدگذاری، بیت‌ها نه تنها برای نمایش اعداد دودویی، بلکه برای سایر سمبل‌های گسسته، همچون ارقام ده‌دهی و یا حروف الفبا نیز به کار برده می‌شوند. با استفاده صحیح از مجموعه‌های دودویی و به کارگیری روش‌های مختلف کدگذاری، می‌توان گروه‌های بیت‌ها را برای ساخت مجموعه‌های کامل دستورالعمل‌ها جهت انجام محاسبات به کار برد.

۳.۴.۱ کدگذاری DNA

همان‌گونه که اشاره شد، در علوم اطلاعات، کدگذاری دیجیتالی دودویی بنیادی‌ترین روش کدگذاری است که هر نوع اطلاعاتی را می‌توان به‌وسیله دو حالت ۰ یا ۱ یا ترکیبی از آنها، کدگذاری کرد. در رشته DNA چهار نوع باز آلی، آدنین (A) و تیمین (T) و سیتوزین (C) و گوانین (G) وجود دارند که ساده‌ترین کدگذاری متناظر با کدنویسی چهار باز نوکلئوتید (A، T، C، G)، استفاده از چهار رقم ۰(۰۰) و ۱(۰۱) و ۲(۱۰) و ۳(۱۱) می‌باشد. به وضوح $۲۴ = !۴$ الگوی کدنویسی با این قالب کدگذاری وجود دارد. همچنین ما می‌دانیم مطابق با قاعده مکملی واتسون - کریک، در یک رشته دوگانه از DNA دو زنجیره به وسیله پیوندهای هیدروژنی مابین زوج بازها به یکدیگر متصل شده‌اند، به‌طوری که همیشه آدنین با تیمین (دو پیوند هیدروژنی) و گوانین با سیتوزین (سه پیوندهیدروژنی) پیوند برقرار می‌کند بنابراین برای رعایت مشخصات بیولوژیکی ۴ باز نوکلئوتید، قانون مکملی $(۱ = ۰ \sim)$ و $(۰ = ۱ \sim)$ در کدگذاری دیجیتال پیشنهاد می شود. بنابراین مطابق با این قانون ۰(۰۰) با ۳(۱۱) و ۱(۰۱) با ۲(۱۰) مکمل است. لذا از میان این ۲۴ نوع الگو تنها ۸ نوع الگوی $۰۱۲۳/CTAG$، $۰۱۲۳/CATG$، $۰۱۲۳/GTAC$، $۰۱۲۳/GATC$، $۰۱۲۳/TCGA$، $۰۱۲۳/TGCA$، $۰۱۲۳/ACGT$ و $۰۱۲۳/AGCT$ ساختاری متناسب با قانون مکملی بازهای نوکلئوتید دارند که در جدول ۱.۱ نشان داده شده است.

کدگذاری دیجیتال دودویی از توالی DNA نسبت به کدگذاری حرفی DNA دارای مزایای زیر است:

۱: کاهش افزونگی کدگذاری اطلاعات و بهبود بهره‌وری کدگذاری، در مقایسه با کدگذاری حرفی DNA در

[۱] Binary
[۲] Bit