آموزش

رمزنگاری و رمزگشایی

سید محمدرضا هاشمی موسوی
hashemi_moosavi@yahoo.com

اشاره

در قسمت قبل با مقدمه‌ای از رمزنگاری‌های متعارف در قرون گذشته به‌صورت توصیفی آشنا شدید. در این قسمت و قسمت‌های بعدی با اصول رمزنگاری و رمزگشایی‌های «تک‌الفبایی حاصل از الفبای متعارف مستقیم»، «رمزهای تک‌الفبایی مبتنی بر تبدیل‌های خطی» و «سیستم‌های چندحرفی» و... آشنا خواهید شد. در این قسمت به رمزهای تک‌الفبایی حاصل از الفبای متعارف مستقیم (رمز جای‌گذاری) می‌پردازیم. در آخر نیز برای آزمودن یافته‌ها، چند تمرین خواهیم آورد.

کلیدواژه‌ها:
پیام رمز، رمز جای‌گذاری، دنبالهٔ صریح، دنبالهٔ رمزی، رمزهای تک‌الفبایی، الفبای جای‌گذاری، عدد کلیدی، کلید رمز، فراوانی حروف.

گشودن الفباهای متعارفِ مستقیم از راه تکمیل دنبالهٔ صریح

اکنون که می‌دانیم چگونه می‌توان پیامی را با استفاده از یک الفبای متعارف مستقیم به رمز درآورد، گشایش رمز پیامی را که از این راه به رمز درآمده است، بررسی می‌کنیم. برای مثال، پیام زیر را در نظر می‌گیریم:

BPM VMOWBQIBQWVA NWZ I AMBBTMUMVB WN BPM ABZQSM IZM IB IV QUXIAAM ZMKWUUMVL EM QVKZMIAM WCZ WNNMZ

حال خود را در موقعیت یک رمزگشا قرار می‌دهیم که نسخه‌ای از این پیام را به‌دست آورده است و می‌خواهد آن را بگشاید. به‌علاوه، فرض می‌کنیم که رمزگشا به طریقی (شاید با حدسی اتفاقی یا به‌علت آشنایی با روش‌های رمزنگاری فرستندهٔ پیام) سیستم کلی را می‌داند، اما از کلید ویژه بی‌اطلاع است. در چنین موقعیتی، تنها کاری که او باید انجام دهد، یافتن عددی است که وی را به خواندن پیام قادر سازد؛ عددی که نشان‌دهندهٔ میزان انتقال دنبالهٔ رمزی نسبت‌به دنبالهٔ صریح باشد. از این نظر، مسئله زیاد مشکل به نظر نمی‌رسد. در کل تنها ۲۵ عدد ممکن وجود دارند و هر کدام را می‌توان به‌نوبت امتحان کرد تا عددی به دست آید که پیام را فاش سازد. درواقع، ممکن است کار کردن با یک کلمه برای یافتن کلید ویژه کافی باشد.

بنابراین اگر برای مثال کلمهٔ اول پیام «BPM» را در نظر بگیریم، معادل‌های عددی حرف‌های این کلمه ۲، ۱۶ و ۱۳ هستند. اکنون اگر عدد ۱ را از هریک از این عددها کم کنیم:

۱	۱۵	۱۲
↕	↕	↕
A	O	L

و اگر عدد ۲ را از هریک از عددهای نخست کم کنیم:

۲۶	۱۴	۱۱
↕	↕	↕
Z	N	K

این فرایند را ادامه می‌دهیم و نتیجه را به‌صورت یک جدول منظم می‌کنیم. از ۲، ۱۶ و ۱۳ شروع می‌کنیم و اطلاعاتی را که به‌دست می‌آوریم به‌ترتیب زیر جدول‌بندی می‌کنیم:

متناظر حرفی		
A	O	L
⋮	⋮	⋮
T	H	E

اعداد حاصل		
۱	۱۵	۱۲
⋮	⋮	⋮
۲۰	۸	۵

مقدار کم‌شده
۱
⋮
۸

وقتی به عدد ۸ می‌رسیم، کلمهٔ THE را که مناسب به نظر می‌رسد مشاهده می‌کنیم و درواقع، اگر کلید ۸ را برای تمام پیام به‌کار ببریم، می‌توانیم تمام پیام را بخوانیم.

تذکر: خواننده باید روی جزئیات از رمز درآوردن بسیار تمرین کند تا بر فرایند رمزگشایی مسلط شود.

توجه کنید که اگر عمل کم کردن را با ظهور کلمهٔ THE متوقف نکرده بودیم و کار را تا آخرین عدد، یعنی ۲۵ ادامه می‌دادیم، هریک از سه ستون سمت راست جدول، یک الفبای کامل اما به‌ترتیب معکوس می‌بود. در صورت تمایل، می‌توان این عددهای کلیدی را به‌ترتیب معکوس، امتحان کرد. یعنی اول ۲۵، بعد ۲۴، تا به آخر و درنهایت در هر ستون، الفباها به‌ترتیب مستقیم قرار گیرند. چنین وضعی، روشی را که اندکی با روش فوق تفاوت دارد برای جست‌وجوی کلید به‌دست می‌دهد.

تمرین: پیام‌های زیر را بگشایید.

1. VXMDUJA JARCQVNCRL LXDUM KN LXWBRMNANM ANVJRWMNA JARCQVNCRL.

2. MZVYDIB DN DJ OCZ HDIY RCVO ZSZMXDNZ DN OJ OCZ WJYT.

(راهنمایی: برای حل تمرین‌ها: **۱.** عدد کلیدی K برابر ۹ است: K=۹
۲. عدد کلیدی K برابر ۲۱ است: K=۲۱.)

گشودن الفباهای متعارف مستقیم با استفاده از فراوانی حرف‌ها

در این‌جا به توضیح روش مبتنی بر انتقال الفبای معمولی می‌پردازیم. این روش بر یک ویژگی اساسی زبان استوار است:
«فراوانی نسبی به کار رفتن حرف‌های مختلف الفبا»

یک نمونه از متنی به زبان صریح را انتخاب می‌کنیم. برای مثال، صفحه‌ای از یک کتاب یا چند سطر از یک روزنامه (به زبان لاتین) را انتخاب می‌کنیم و فراوانی هر حرف را می‌شماریم؛ یعنی معلوم می‌کنیم که هر حرف چند بار به کار رفته است. برای مثال از نمونه‌ای که به طول ۱۰۰۰ حرف انتخاب کرده‌ایم، پس از شمارش، نتیجهٔ زیر به‌دست آمده است:

A	۷۳	J	۲	S	۶۳
B	۹	K	۳	T	۹۳
C	۳۰	L	۳۵	U	۲۷
D	۴۴	M	۲۵	V	۱۳
E	۱۳۰	N	۷۸	W	۱۶
F	۲۸	O	۷۴	X	۵
G	۱۶	P	۲۷	Y	۱۹
H	۳۵	Q	۳	Z	۱
I	۷۴	R	۷۷		

فراوانی نسبی هر حرف، که مورد نظر ماست، درصد تعداد دفعات ظاهر شدن آن حرف، یعنی درصد تعداد دفعات به کار رفتن آن حرف است. از آن‌جا که تعداد کل حروف در این نمونه ۱۰۰۰ است، فراوانی نسبی هر حرف از تقسیم فراوانی واقعی آن بر ۱۰ به‌دست می‌آید.

همان‌طور که انتظار می‌رود، فراوانی نسبی حروف مختلف فرق دارد. تعداد **Eها** ۱۳٪ کل حرف‌ها، تعداد Tها تقریباً ۹٪ کل حرف‌ها و تعداد هریک از حروف صدادار A، I و O حدود **۷**٪ کل حرف‌هاست. بعضی حروف مانند G، V، W و Y به‌ندرت به‌کار رفته‌اند (تعداد آن‌ها بین ۱ تا ۲ درصد بوده است) و حرف‌های J، K، Q و Z تقریباً اصلاً به‌کار نرفته‌اند.

بیان این‌که «فراوانی نسبی E **۱۳**٪ است» به این معناست که در یک انتخاب تصادفی از کل هزار حرف، احتمال به‌دست آمدن یک E به نسبت ۱۳ به ۱۰۰ است. از این‌رو، اصطلاح احتمال را که تعریف دقیق ریاضی آن مبتنی بر مفهوم فراوانی نسبی است، به‌کار می‌بریم و از نماد $P_E=۰/۱۳$ استفاده می‌کنیم تا نشان دهیم که احتمال به‌دست آوردن یک E برابر **۱۳**٪ است. با این نمادگذاری:

$$P_A=۰/۰۷ \quad و \quad P_T=۰/۰۹ \quad و \quad ... \quad و \quad P_Z\approx ۰$$

از آن‌جا که مجموع همهٔ فراوانی‌ها برابر تعداد همهٔ حرف‌ها در نمونه است، نتیجه می‌شود که مجموع فراوانی‌های نسبی ۱۰۰ درصد، یعنی ۱ است و بنابراین مجموع احتمال برای همهٔ ۲۶ حرف برابر ۱ است:

$$P_A+P_B+P_C+\ldots+P_Z=۱$$

این عبارت را به‌صورت اختصاری زیر نشان می‌دهند:

$$\sum_{i=A}^{i=Z} P_i = ۱$$

عبارت سمت چپ آن چنین خوانده می‌شود:

«مجموع مقادیر P_i وقتی که i به‌نوبت مقادیر از A تا Z را اختیار می‌کند.»

A، حد پایین و Z، حد بالای مجموع نامیده می‌شود.

هر متنی از زبان متداول را که به اندازهٔ کافی طولانی باشد، بررسی کنیم، نتیجه‌ای مشابه با آنچه از نمونهٔ هزار حرفی به‌دست آوردیم، به‌دست خواهیم آورد.

درصد تعداد دفعات به‌کار رفتن هر حرفی در نمونه‌های طولانی را «**فراوانی**» مشخصهٔ آن حرف می‌نامند. اگر پیامی بسیار کوتاه باشد، ممکن است فراوانی نسبی بعضی از حروف در آن پیام با فراوانی مشخصهٔ آن‌ها تفاوت زیادی داشته باشد. اما هرچه پیام طولانی‌تر باشد، احتمال کمتری وجود دارد که تفاوت‌های زیادی با فراوانی‌های مشخصه وجود داشته باشد.

در این‌جا اطلاعات مربوط به شمارش فراوانی‌ها را که یادداشت کردیم به شکل یک نمودار میله‌ای نمایش می‌دهیم. برای سادگی، مقدار هریک از فراوانی‌های نسبی را به نزدیک‌ترین عدد صحیح گرد می‌کنیم، به گونه‌ای که بتوان فراوانی مشخصهٔ هر حرف را مانند زیر با نشان خط‌هایی نمایش داد:

(الگوی فراوانی مشخصهٔ هر حرف در دنبالهٔ صریح)

A , B , C , D , E , I , ... , X , Y , Z

در این نوع نمایش، نموداری می‌بینیم که واضح‌تر از جدول ارقام، زیاد و کم بودن فراوانی مشخصهٔ حرف‌ها را آشکار می‌سازد.

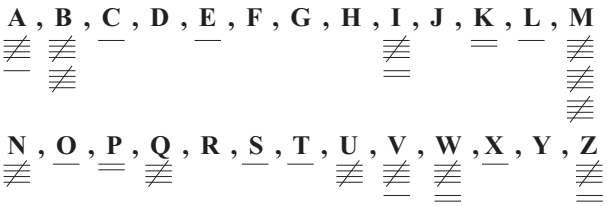
در این نمودار یک الگوی با اهمیت وجود دارد. در بین حروفی که فراوانی‌های زیاد دارند، A، E و I را می‌بینیم که به فاصله‌های برابر (سه حرف درمیان) قرار دارند و E بیشترین فراوانی را دارد. هم‌چنین زوج متوالی N و O و نیز سه‌تایی R، S و T را می‌بینیم. در بین حروفی که فراوانی‌های کم دارند، زوج متوالی J و K و دنبالهٔ U، V، W، X، Y و Z قرار دارند.

این الگوی فراوانی‌های زیاد و کم با فاصله‌های مذکور، مشخصهٔ الفبای معمولی در زبان صریح و یک ابزار اساسی رمزگشایی است.

حال ببینیم هنگامی که پیامی از راه انتقال الفبای معمولی نسبت به خود الفبا به رمز درآید، چه اتفاقی می‌افتد. برای مثال، فرض کنید میزان انتقال، هشت حرف باشد. پس هربار که هر حرف A در پیام صریح ظاهر می‌شود به جای آن I گذاشته می‌شود. برای

نمایش نمادی این جای‌گذاری، قرار می‌گذاریم که حروف زبان صریح را با اندیس P و حروف رمزی را با اندیس C مشخص کنیم. با این قرارداد، I از زبان رمزی به جای A از زبان صریح را به‌صورت $A_P=I_C$ نشان می‌دهیم. هم‌چنین هربار که B در پیام ظاهر شود، به جای آن J گذاشته خواهد شد، یعنی $B_P=J_C$. به جای هر حرف از پیام اصلی، هرجا که ظاهر شود، همیشه یک حرف که معادل آن است می‌آید. نتیجه آن است که نمودار توزیع فراوانی پیام رمزی همان نمودار توزیع فراوانی پیام صریح اولیه خواهد بود، با این تفاوت که به اندازهٔ ۸ مکان انتقال یافته است. اگر کلید ویژه عدد دیگری مانند n باشد، (نمودارِ) توزیع نیز به اندازهٔ n مکان انتقال می‌یابد.

اکنون پیام رمزی را در نظر می‌گیریم که توزیع زیر برای آن به دست آمده است:



این نمودار به دلیل نشان دادن فراوانی تک‌تک حروف «توزیع فراوانی تک‌حرفی» نامیده می‌شود. حال سعی می‌کنیم الگوی الفبای معمولی را در این توزیع بیابیم. مشاهده می‌کنیم که I، M و Q (همه با فراوانی زیاد) سه حرف در میان هستند و M بیشترین فراوانی را دارد؛ V و W زوجی از حروف متوالی با فراوانی‌های زیاد است؛ Z، A و B یک سه‌تایی از حروف متوالی با فراوانی‌های زیاد است؛ C، D، E، F، G و H، دنباله‌ای طولانی از حروف متوالی با فراوانی‌های کم را تشکیل می‌دهند.

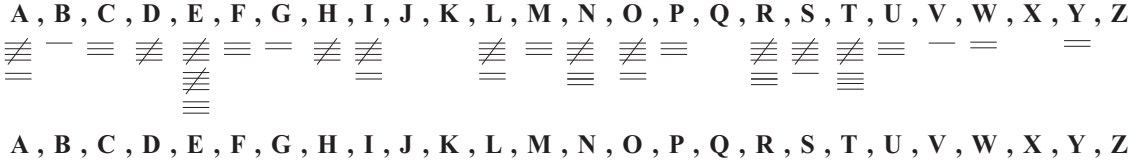
اگر قرار دهیم $I_C=A_P$، آن‌گاه تمام این مطالب با وضعیت یک الفبای متعارف مستقیم جور درمی‌آیند. بنابراین، حرف I در این توزیع باید متناظر حرف A در الفبای معمولی باشد.

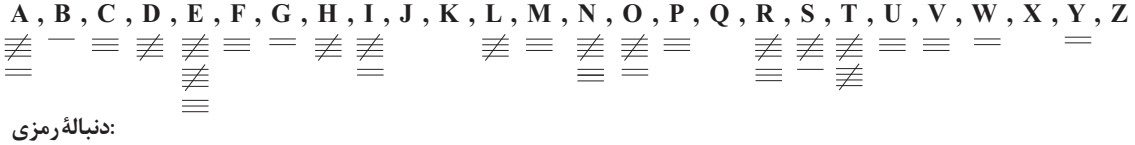
روش دیگر مشاهدهٔ این مطلب، لغزاندن توزیع مربوط به رمز در مقابل توزیع مربوط به زبان صریح به گونه‌ای است که آن‌ها با یکدیگر متناظر شوند.

ابتدا توزیع زبان صریح را می‌نویسیم و برای آن‌که عمل متناظر کردن ممکن باشد، آن را دو بار به‌دنبال هم می‌نویسیم. آن‌گاه توزیع رمز را مقابل آن قرار می‌دهیم و هر بار به اندازهٔ یک حرف آن را می‌لغزانیم تا یک تناظر خوب از فراوانی‌ها، یعنی زیادها در مقابل زیادها و کم‌ها در مقابل کم‌ها به‌دست آید.

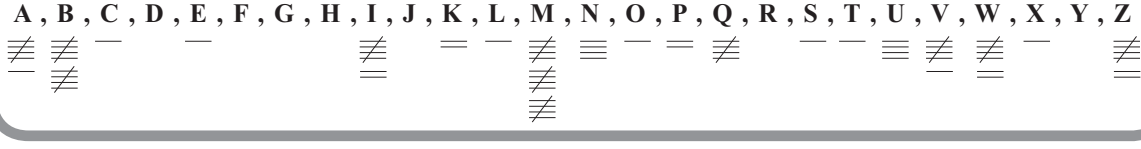
اگر این دو توزیع را در دو صفحهٔ متفاوت قرار دهیم، آن‌گاه لغزاندن یکی در مقابل دیگری ساده‌تر خواهد شد. کار را با قرار دادن A_C در مقابل B_P شروع می‌کنیم (شکل ۳). در این وضعیت در تعداد کمی از مکان‌ها می‌بینیم که فراوانی‌های زیاد متناظر با فراوانی‌های زیاد قرار گرفته‌اند، برای مثال، N_P و O_P، M_C و N_C و R_P، Q_C و A_P و Z_C؛ و نیز در تعداد کمی از مکان‌ها فراوانی‌های کم‌متناظر یکدیگر قرار گرفته‌اند، برای مثال، K_P و J_C، Y_P، X_C و Z_P و Y_C. اما موارد مهمی از ناسازگاری نیز وجود دارد. یعنی E_P، که بیشترین فراوانی را دارد، مقابل حرفی است که به‌هیچ‌وجه ظاهر نمی‌شود و سازگاری فراوانی‌ها در زوج‌های B_P و A_C، I_P و H_C، S_P و R_C، T_P و S_C، X_P و W_C، بسیار ضعیف است. نتیجه می‌گیریم که این تناظر مناسب نیست.

دنبالهٔ صریح:





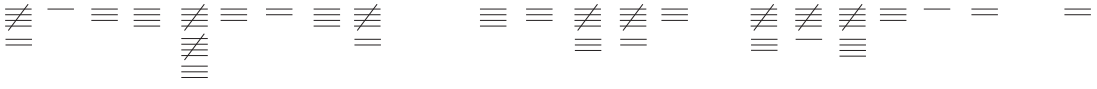
دنبالهٔ رمزی:



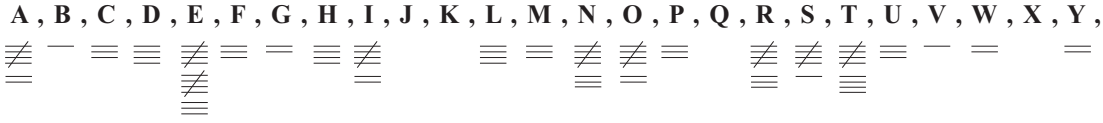
حال توزیع رمز را چنان انتقال می‌دهیم که $A_C=C_P$ و دو توزیع را مقایسه می‌کنیم (شکل ۴). دوباره نتیجه می‌گیریم که تناظر خوبی نداریم.

دنبالهٔ صریح:

A , B , C , D , E , F , G , H , I , J , K , L , M , N , O , P , Q , R , S , T , U , V , W , X , Y , Z

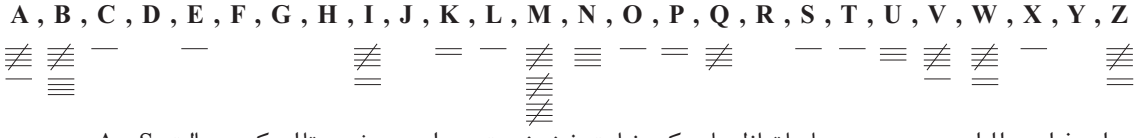


A , B , C , D , E , F , G , H , I , J , K , L , M , N , O , P , Q , R , S , T , U , V , W , X , Y , Z



دنبالهٔ رمزی:

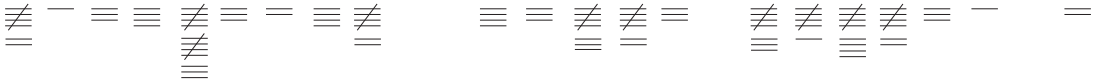
A , B , C , D , E , F , G , H , I , J , K , L , M , N , O , P , Q , R , S , T , U , V , W , X , Y , Z



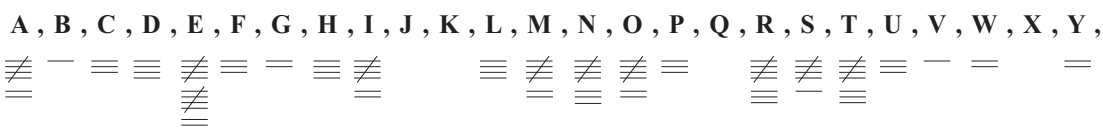
این فرایند را ادامه می‌دهیم و هر بار با تناظرهایی که رضایت‌بخش نیستند مواجه می‌شویم تا این‌که به حالت $A_C=S_P$ می‌رسیم (شکل ۵).

دنبالهٔ صریح:

A , B , C , D , E , F , G , H , I , J , K , L , M , N , O , P , Q , R , S , T , U , V , W , X , Y , Z

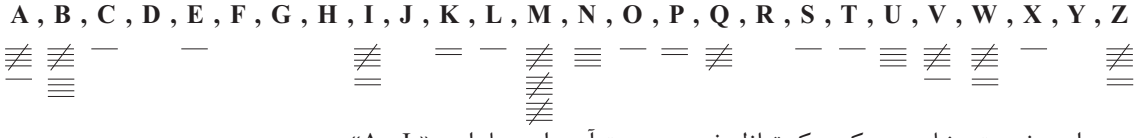


A , B , C , D , E , F , G , H , I , J , K , L , M , N , O , P , Q , R , S , T , U , V , W , X , Y , Z



دنبالهٔ رمزی:

A , B , C , D , E , F , G , H , I , J , K , L , M , N , O , P , Q , R , S , T , U , V , W , X , Y , Z



در این وضعیت مشاهده می‌کنیم که تناظر خوبی به‌دست آورده‌ایم، بنابراین: «$A_P=I_C$».

تمرین:

پیام‌های زیر را با متناظر قرار دادن توزیع آن‌ها با توزیع فراوانی زبان صریح بگشایید.

۱.

CQSOB KOHSF WG PZIS PSQOIGS RWFH DOFHWQZSG WB HVS KOHSF FSTZSQH GIBZWUVH PIH HVS KOHSF OPGCFPG FSR OBR MSZZCK HVS UFSSBG OBR PZISG HVOH OFS ZSTH AOYS HVS RSSD PZIS CQSOB.

(پاسخ ۱. عدد کلیدی، یعنی تعداد حروف انتقال‌یافته، برابر ۱۴ است: K=۱۴)

۲.

SNHPJQ NX F MJFAD XNQAJW BMNYJ RJYFQQNH JQJRJSY NY NX RFLSJYNH YFPJX F MNLM UTQNXM FSI ITJX STY YFWSNXM TW WZXY JFXNQD.

(پاسخ ۲. عدد کلیدی، یعنی تعداد حروف انتقال‌یافته، برابر ۵ است: K=۵)