

به نام خدا

مثال - منبع  $X$  اطلاعاتی را از مجموعه  $X = \{a, b, c, d\}$  با احتمالات  $\{1/2, 1/4, 1/8, 1/8\}$  تولید می کند. می خواهیم اطلاعات این منبع را به شکل مناسب که کمترین (کوچکترین) منبع) در روی کانال فرستیم درش مناسب را پیدا کنیم (که اینجاست)

$$X = x_1 = a$$

$$Pr \{X = x_1\} = \frac{1}{2}$$

$$M^* = \{0, 1\}$$

$$X = x_2 = b$$

$$Pr \{X = x_2\} = \frac{1}{4}$$

$$X = x_3 = c$$

$$Pr \{X = x_3\} = \frac{1}{8}$$

$$X = x_4 = d$$

$$Pr \{X = x_4\} = \frac{1}{8}$$

ادش پشتهای 1) در درشت پشتهای اول برای حرکت از خودی های منبع یک مکده که به طول 2 در نظر گرفته

$$C(x_1) = 55$$

$$C(x_2) = 51$$

$$C(x_3) = 15$$

$$C(x_4) = 11$$

که منبع با طول ثابت  $l = 2$  (که  $C_1$ )

$$L(C) = E\{l(x)\} = 2$$

$$H(X) = \frac{7}{4}$$

همان طور که دیدیم  $L(C) > H(X)$  برقرار است و با این روش نتوانستیم به کمترین میانگین

طول که برابر  $H(X)$  برسم. اما پیاده سازی این که در سمت اندر در یک درگاه است و برابر با یک

احداث نیز بدون ابهام امکان پذیر است. (باله منبع با طول ثابت یعنی توان به کمترین میانگین طول کلمات که رسد)

(تقریباً خواصی از توزیع آماری منبع)

در رافع به دلیل آنکه در حراحی که منبع به خصوصیات آماری منبع (احتمال انفجاری خودی منبع)  
ترجمه نکردیم، نتوانستیم به کمترین میانگین طول کلمات گذر کنیم.  
برای آنکه بتوانیم به بهترین شکل از پهنای ایند کانال استفاده کنیم، راضانات لکتری را روی کانال ارسال  
کنیم، لازم است که در حراحی که منبع، به خصوصیات آماری منبع نیز توجه داشته باشیم.

« برای همین به کمترین میانگین طول کلمات که لازم است که طول کلمات که را برای پهنای خودی منبع بپردازیم  
به احتمال آنها، متغیر در نظر بگیریم. به عبارت دیگر برای انفجاری خودی منبع که احتمال وقوع بیشتری  
دارند، کلمه که کوآتری اختصاص می دهیم. به این صورت با احتمال بیشتر، کلمه که کوآتری روی  
کانال فرستاده می شود و از پهنای ایند کانال به شکل لکتری استفاده می شود. افزادگان لکتری روی کانال فرستاده می شود

خدا صبح در کلاس این درس می رانیم به کمر تن میایم طرکات که رستم

درس پیشگامی درم که  $C_2$

$$X = x_1 = a, \quad P_r \{X = x_1\} = \frac{1}{2}$$

$$X = x_2 = b, \quad P_r \{X = x_2\} = \frac{1}{4}$$

$$X = x_3 = c, \quad P_r \{X = x_3\} = \frac{1}{8}$$

$$X = x_4 = d, \quad P_r \{X = x_4\} = \frac{1}{8}$$

$$C(x_1) = 0$$

$$C(x_2) = \underline{10}$$

$$C(x_3) = \underline{101}$$

$$C(x_4) = 111$$

در این حالت، میانگین طول کلمات برابر است با:

$$L(C_2) = E\{l(x)\} = \sum_{i=1}^4 p(x_i) l(x_i)$$

$$= \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{8} \times 3 = \frac{7}{4}$$

از طرف دیگر، سی دانسی  $H(X) = \frac{7}{4}$  بنابراین برای که  $C_2$  در بریم

$$L(C_2) = H(X)$$

یعنی با این روش توانیم به کمترین میانگین طول کلمات که برسیم.

اما به دلیل اینکه طایفه که  $\epsilon$ ، با جدول ثابت نیستند، به جدولی انگرد در منبع، در یک منبع نیست.  
 که  $\epsilon$  نیز است. به خصوص در سمت دیگر که لازم است با دریافت رشته ای از بیت ها،  
 عملیات بازایی اعداد را انجام شود زیرا در یک اید سترانه طایفه که بدون اتمام از رشته بیت  
 دریافتی، استخراج کند و بازایی اعداد را انجام دهد. به طور مثال، فرض کنیم که رشته بیت را در  
 دریافت کرده ایم،

$0101011100\dots$ 
  
 $a$ 
  
 $b$ 
  
 $c$

از یک - با دریافت رشته بیت اطلاعات دریافتی ، باید ابتدا کلمات که از رشته بیت دریافتی استخراج  
گردد (puncturing) سپس با توجه به جدول کلمات که ، از تایی اطلاعات منتج را انجام دهد.  
در که به رشته بیت مثال ، بلکه بعد از مشخص کردن کلمه که اول که برابر '0' است ،  
برای مشخص کردن کلمه که دوم ، در جایگاه سوم می شود . زیرا می تواند صمیم بلکه ده کلمه که  
عربی برابر '۱۵' (یعنی ط) است یا '۱۵۱' (یعنی ح) است که به زبان  
آن عمل های عربی دریافت شده اند . بنابراین در که به تایی اطلاعات بدون ابهام  
امکان پذیر نیست .

ترجیه داشته باشیم که صفای در بازایی امله‌هاست حصاره و محدود دارد طراحی باید به گونه‌ای باشد که این خط  
قابل کنترل در میزان رطوبت کم باشد.

اما اگر طراحی که به درستی انجام نشده باشد و ابهام در بازایی امله‌ها داشته باشیم، لازم است  
که طراحی را تغییر دهیم و ابهام را برطرف کنیم.

در که  $C_2$  به رسل اندک  $C(n_2)$  پیوند  $C(n_3)$  است، در هنگام هراسازی (Puncturing)  
کلمات که از رشته بیت دریافتی، ابهام وجود دارد که باعث ایجاد خطای غیر قابل کنترل می‌شود.

« برای طراحی که دارای منبع اچ‌دی غیر یکنواخت است که کلمات که برای گونه‌ای در نظر بگیریم که صیغ لری

پیشوند کلمه که دیگری نباشد. به این ترتیب، کلمه‌ای ترانه بدون ابهام، طلمات که دارد، نسبت به  
 ریاضی مشخص کند. Self Puncture

روش پیشوندی (سم) که  $C_3$

$$C(x_1) = 0$$

$$C(x_2) = 10$$

$$C(x_3) = 110$$

$$C(x_4) = 111$$

$$L(C_3) = \frac{7}{4} = H(X)$$

به دلیل آنکه هیچ‌کدام از پیشوندها که دیگری نیست،  
 ابهامی در مبداء سازی طلمات که، و از این ابهامات نخواهیم داشت.

به عنوان مثال، رشته بیت درایستی زیرادر نظر می گیریم.

0 1 0 1 0 1 0 0 0 ...  
a b b d a a

به این ترتیب در دیکه، جداسازی طلمات که بدون ابهام انجام می شود، رسیدن به بازاری اطمینان  
منبع انزوری طلمات که درایستی انجام می شود.

ما به توجه داشته باشیم که به دلیل وقوع خطا در حال، ممکن است بیت هایی در رشته بیت درایستی  
با خطا درایست شوند و باعث ایجاد خطای دیکه شوند. انتظار خطا در بازاری اطمینان از رشته  
بیت درایستی شوند. این مسئله را با کمک روش های مناسب که دیکه در دیکه منبع می توان حل کرد.

حقی  $\alpha$  کدیکتیک را آماره ممکن، کم کرد. (خطای - میزان دشوار کم)

در ادامه سی خواص، مطالب گفته شده در مثالهای قبل را با چارچوب ریاضیاتی تر بیان کنیم. خصوصیات لازم برای کدهای منبع را به دست بیاوریم (در حالت  $\alpha$  lossless)

\* اولین خصوصیت یک کد منبع، غیر سینگولر (non-singular) بودن آن است. منظور از غیر سینگولر بودن که، این است که حریف از الفبای منبع  $\alpha$ ، یک کد منحصراً برای خود داشته باشد. با بیان ریاضی می توان گفت،

$$\forall x_1, x_2 \in X \quad ; \quad x_1 \neq x_2 \quad \Rightarrow \quad C(x_1) \neq C(x_2)$$

باینه صِدِّ مَعَادِل (عَكْسِ نَقِیْن)

$$C(x_1) = C(x_2) \quad \Rightarrow \quad x_1 = x_2$$

این شرط، اولین شرط برای این است که بازایی اشیاء در رابطه بودن ابهام قابل انجام باشد.

حالتی که صله نداشتیم، برای اینکه بازایی اعداد از رشته بیت در مینی، بدون ابهام  
 امکان پذیر باشد، دیگر باید بتواند کلمات که را بدون ابهام از رشته بیت در مینی استخراج کند  
 یا بیان دیگر، که صنف  $self\ puncture$  باشد.

اگر برای حل این مسأله پیشتر درجه یک علامت اضافی مانند  $0$  یا  $1$  اضافه به عملهای  
 کلمات که اضافه شود، در واقع یک سر بار  $overhead$  به سیستم اعمال کرده ایم زیرا  
 یک عمل اضافی به یکدیگر  $M^*$  افزودن می شود که از آن به صورت بسته استفاده  
 نمی شود.

$$M^* = \{ \text{0, 1} \} \xrightarrow{\text{اضافه کردن 0 یا 1 بین کلمات}} M^* = \{ \text{00, 01, 10, 11} \}$$

که با این  $M^*$  :

همان طور که در مثال صدای سِل دیدیم، یک راه حل مناسب برای این مسئله، این است که کلی  
که از بگزنای در نظر بگیریم که هیچ کدی پیشوند طه که دیگری نباشد. به این کدها، کدهای  
لحظه‌ای یا کدهای Prefix یا کدهای self puncture می‌گوئیم.  
instantaneous

\* برای اینکه بتوانیم خصوصیت لحظه‌ای بودن که منبع را به صورت ریاضی بیان کنیم، لازم است  
که ابتدا ضاهیم و معادنی اِ مطرح کنیم.

مفهوم سبک که  $C$

منظور از سبک که  $C$ ، این است که طبات که از صورت یک رشته بیت، بیت ترجم  
مترادهم، به عبارت دیگر

طبه که مربوط به  $x_1$

طبه که مربوط به  $x_n$

$$C(x_1, x_2, \dots, x_n) = C(x_1)C(x_2) \dots C(x_n)$$

سبک  $C$  به الگوی  
 $x_1, \dots, x_n$

بیت ترجم شدن - Concatenation طبات که  $C(x_1), \dots, C(x_n)$

در واقع سبک که  $C$ ، یک تصویر سازی از دنباله های به هم ملحق که در از الگوی منبع  $X$  به دنباله های

باجل محدود از  $M^*$  است.

$$x_1, x_2, \dots, x_n \in X$$

به عنوان مثال: اگر  $C(x_1) = 0$  و  $C(x_2) = 10$  آنجا

$$C(x_1, x_2) = 010$$

مفهوم 'به طور یکتا قابل رد کردن' یک کد  
Uniquely Decodable  
یک کد  $C$  به طور یکتا قابل رد کردن است اگر هیچ کد دیگری  
نیست که هر رشته از اشیای منبع به یک رشته مفرد از کلمات که صدق می شود  
به این ترتیب با دربارین هر رشته بیت از کلمات که می توان به رشته از اشیای منبع به صورت مفرد  
قرار، بازایی کرد

$$C(x_1, x_2) \longrightarrow x_1 x_2$$

$$C(x_1, x_2, x_3) \longrightarrow x_1 x_2 x_3$$

تکته ای که لازم است به آن توجه کنیم، این است که در این کدها، برای اینکه عملیات دیکریند انجام دهیم، لازم است که رشته بیت اطلاعات به خودی خود کامل دریافت شده باشد. اما در عمل ما علامه مند هستیم که عملیات دیکریند را با شروع از هر لحظه از رشته بیت درایستی تا پایان در هر لحظه از رشته بیت درایستی، بتوانیم بدون انجام انجام دهیم. به همین دلیل مفهوم لحظه ای بودن که منع احصیت پیدا می کند.

مفهوم کدهای یایکه Prefix یا که Self puncture

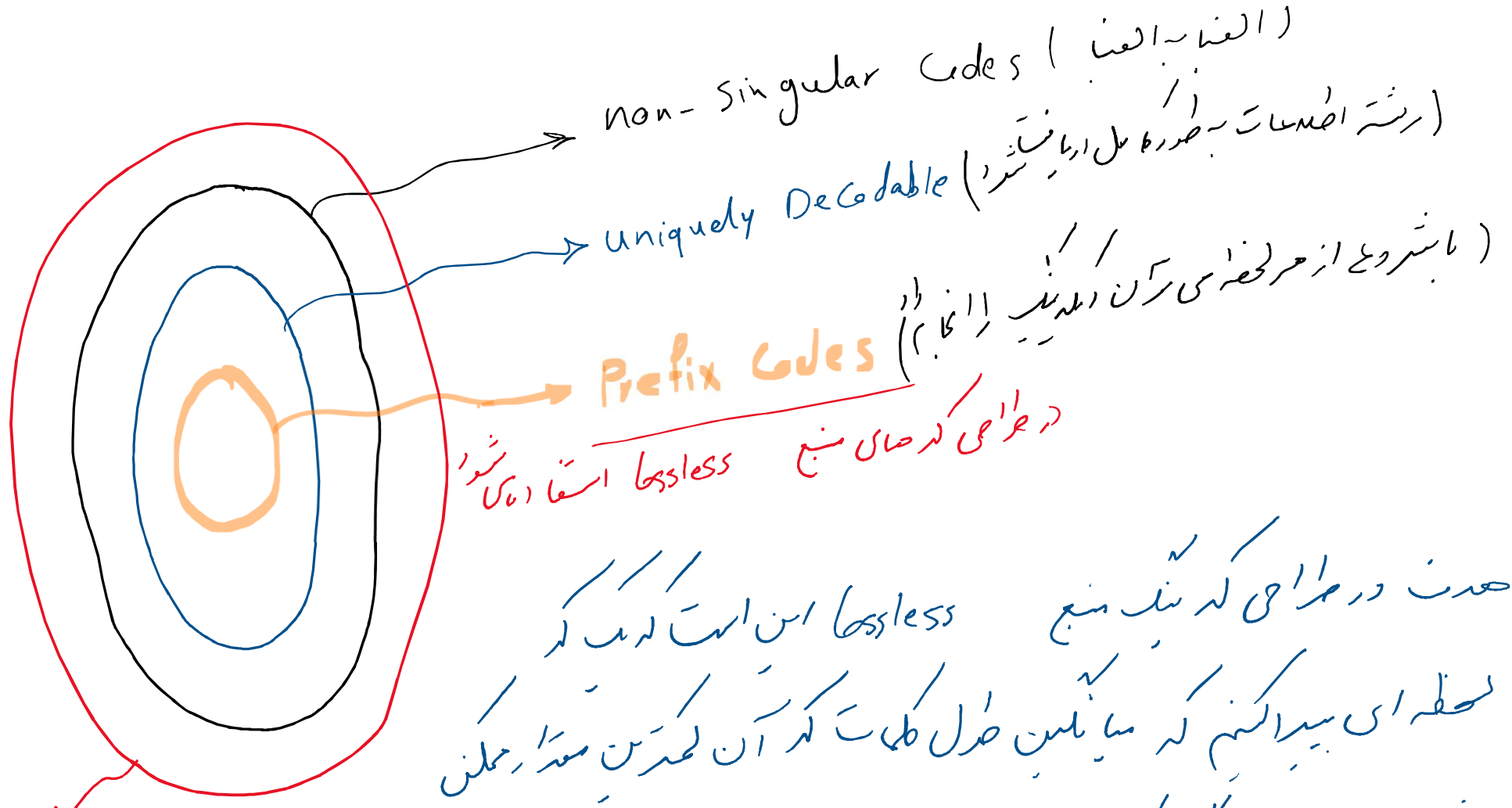
اگر دیک که منبع، صیغ کدهای پیشوند کدهای دیگری نباشد، کدهای یایکه Prefix  
میگیریم. در این صورت با شروع از هر کدهای از رشته بیت در این می توانیم Puncturing  
را بدون ابهام انجام دهیم و اطمینان منبع را با این کنیم.

به عنوان مثال

| کدهای یایکه | الفبای منبع |
|-------------|-------------|
| 0           | a           |
| 10          | b           |
| 110         | c           |
| 111         | d           |

0, 10, 111, 110, 0, 10, ...

a b d c a b



non-singular codes (الفنا به الفنا)

Uniquely Decodable (رشته‌های صدها به صورتی که قابل افتراق باشند)

Prefix Codes (با شروع از هر لحظه‌ای رکن الگوریتم را می‌توان به پایان رسانید)

در طراحی که صای منبع lossless استفا دهی شود

حداث در طراحی که منبع lossless این است که در هر لحظه ای پیدا کنیم که میانگین طول کلمات که آن لحظه در دسترس است ممکن باشد یا به عبارت دیگر کمترین فاصله را با آن نزدیکی منبع داشته باشد

all code  
 $L \geq H(x)$   
میانگین طول کلمات

آنزوی منبع

برای آنکه با طراحی کدهای منبع lossless، آنتروپی، در ادامه بحث در کدهای منبع کاربرد آماری  
 ضوابطی کرد. اولین کد، کد نزدیک به نام کد Shannon-Fano است.  
 دومین کد، کد هانی Huffman است.

Shannon - Fano

★

همان کد که گفتیم، هدف طراحی، این است که یک کد که لحظه‌ای به این منبع که می‌بینیم طول کلمات که آن به  
 آنتروپی منبع نزدیک باشد. می‌دانیم که همواره

$$L \geq H(X)$$

از طرف دیگر داریم،

$$L = E\{l(x)\} = \sum_x \underbrace{p(x)} \underbrace{l(x)}$$

$$H(X) = E\left\{\log \frac{1}{p(x)}\right\} = \sum_x \underbrace{p(x)} \underbrace{\log \frac{1}{p(x)}}$$

با مقایسه روابط بالا، می‌توان نتیجه گرفت که اگر برای الفبای  $x$  از منبع، طول کدها مشخصه  $l(x)$  را برابر  $\log \frac{1}{p(x)}$  در نظر بگیریم به کمترین میانگین طول کلمات که برابر آنزودی منبع است، می‌رسیم. اما باید توجه داشته باشیم که  $l(x)$  یک عدد صحیح است در صورتی که

$\log \frac{1}{P(x)}$  نزوداً یک عدد صحیح است.

در  $\log$  Shannon-Fano که یک عدد نزدیک به  $\log \frac{1}{P(x)}$  است، طول کد که متناظر با الفبای  $x$  از

منبع برابر

$$l(x) = \left\lceil \log \frac{1}{P(x)} \right\rceil$$

در نظر گرفته می شود. (زیرا شرط  $L \geq H(x)$  برقرار باشد تا بازایی اعداد اعشاری به میزان دلخواه کم، امکان پذیر شود)

اگر  $M$ -ary عددی باشد، داریم

$$l(x) = \left\lceil \log_M \frac{1}{P(x)} \right\rceil$$

$$M^* = \{0, 1, \dots, M-1\}$$