

Capítulo 6 Teoría de la Información

Seguridad Informática y Criptografía



Material Docente de
Libre Distribución

Ultima actualización del archivo: 01/03/06
Este archivo tiene: 59 diapositivas

Dr. Jorge Ramíó Aguirre
Universidad Politécnica de Madrid

Este archivo forma parte de un curso completo sobre Seguridad Informática y Criptografía. Se autoriza el uso, reproducción en computador y su impresión en papel, sólo con fines docentes y/o personales, respetando los créditos del autor. Queda prohibida su comercialización, excepto la edición en venta en el Departamento de Publicaciones de la Escuela Universitaria de Informática de la Universidad Politécnica de Madrid, España.

Curso de Seguridad Informática y Criptografía © JRA

Fundamentos de la Seguridad Informática

Los pilares sobre los que descansa toda la teoría asociada a los criptosistemas son básicamente tres:

- La teoría de la información
 - Estudio de la cantidad de información contenida en los mensajes y claves, así como su entropía.
- La teoría de los números
 - Estudio de las matemáticas discretas y cuerpos finitos que permiten las operaciones de cifrado y descifrado.
- La teoría de la complejidad de los algoritmos
 - Estudio de la clasificación de los problemas como computacionalmente tratables o intratables.

Estos temas los veremos en éste y en los siguientes capítulos del libro.

•

•


•

Capítulo 6: Teoría de la Información

Página 180

Teoría de la información

- Definición de información:
 - Es el conjunto de datos o mensajes inteligibles creados con un lenguaje de representación y que debemos proteger ante las amenazas del entorno, durante su transmisión o almacenamiento, usando técnicas criptográficas entre otras herramientas.
 - La teoría de la información mide la *cantidad de información* que contiene un mensaje a través del número medio de bits necesario para codificar todos los posibles mensajes con un *codificador óptimo*.



¿Qué significa cantidad de información y codificador óptimo?

→

© Jorge Ramío Aguirre

Madrid (España) 2006

• • • • •

•

•

•

Capítulo 6: Teoría de la Información

Página 181

Representación de la información

Puede ser numérica, alfabética, simbólica, por lenguaje.

Ejemplo: 15/01/05 15-01-05 15-1-05 15/01/2005
 01/15/05 01-15-05 1-15-05 01-15-2005 ...

- Todos son el día 15 de enero del año 2005.

Vitaminas: B₁₂, C, ...

Grupo sanguíneo: A2 Rh+ ...

Elementos: Fe, Si, Hg ...

Compuestos químicos: H₂O, CO₂ ...

Más común ➡ Lenguaje con código: “¿Hace calor allí?”

Veamos la información que contiene el mensaje
¿Hace calor allí?

© Jorge Ramío Aguirre

Madrid (España) 2006

• • • • •

•

•

•

Capítulo 6: Teoría de la Información

Página 182

La información que tiene un mensaje

Veremos qué información nos entrega un mensaje dependiendo del contexto en que nos encontremos. Esto puede analizarse:

- a) En función de la *extensión* del mensaje recibido.
- b) En función de la *utilidad* del mensaje recibido.
- c) En función de la *sorpresa* del mensaje recibido.
- d) Dependiendo del *entorno* de esa sorpresa.
- e) En función de la *probabilidad* de recibir un mensaje.

Este último enfoque orientado a la ingeniería y usado por Claude Shannon en su estudio es el que aquí nos interesa.

http://es.wikipedia.org/wiki/Claude_E._Shannon

© Jorge Ramío Aguirre

Madrid (España) 2006

• • • • •

•

•

•

Capítulo 6: Teoría de la Información

Página 183

Cantidad de información (caso 1)

En función de la extensión del mensaje

- Ante una pregunta cualquiera, una respuesta concreta y *extensa* nos entregará mayor información sobre el tema en particular, y diremos que estamos ante una mayor “cantidad de información”.
- Pregunta: ¿Hace calor allí? (una playa en particular)
 - Respuesta 1: Sí, hace mucho calor.
 - Respuesta 2: Cuando no sopla el viento, el calor allí es inaguantable pues supera los 42 grados a la sombra. ☹

¿Dónde hay una mayor cantidad de información?

© Jorge Ramío Aguirre

Madrid (España) 2006

• • • • •

Capítulo 6: Teoría de la Información Página 184

Cantidad de información (caso 2)

En función de la utilidad del mensaje

- Ante una pregunta cualquiera, una respuesta más *útil* y clara nos dejará con la sensación de haber recibido una mayor “cantidad de información”.
- Pregunta: ¿Hace calor allí? *(una playa en particular)*
 - Respuesta 1: Sí, sobre 30 grados. ☝
 - Respuesta 2: Si no hay viento del sur y el mar está en calma, es normal que la temperatura suba bastante.

↩ ¿Dónde hay una mayor cantidad de información?

© Jorge Ramío Aguirre Madrid (España) 2006

Capítulo 6: Teoría de la Información Página 185

Cantidad de información (caso 3)

En función de la sorpresa del mensaje

- Ante una pregunta cualquiera, una respuesta más *inesperada* y sorprendente, nos dará la sensación de contener una mayor “cantidad de información”.
- Pregunta: ¿Hace calor allí? *(ahora Finlandia en otoño)*
 - Respuesta 1: Sí, muchísimo. Es insoportable. ☺
 - Respuesta 2: En esta época del año, la temperatura es más suave y el tiempo muy agradable.

↩ ¿Dónde hay una mayor cantidad de información?

© Jorge Ramío Aguirre Madrid (España) 2006

Cantidad de información (caso 4)

Dependencia del entorno (sorpresa)

- Ante una pregunta cualquiera, una respuesta inesperada y *sorprendente* en el entorno, nos dará la sensación de contener una mayor “cantidad de información”.

- **Pregunta: ¿Hace calor allí?**

(ahora las mismas respuestas hablan de la temperatura en un horno)

- Respuesta 1: Sí, muchísimo. Es insoportable.
- Respuesta 2: En esta época del año, la temperatura es más suave y el tiempo muy agradable. ☺?



¿Dónde hay una mayor cantidad de información?

Cantidad de información (caso 5)

En función de la probabilidad de recibir un mensaje

- Este enfoque *probabilístico* es el que nos interesará en cuanto a la definición de Cantidad de Información.

¿Dónde le da alegría a su cuerpo Macarena?

- Respuesta 1: En un país de Europa.
- Respuesta 2: En una ciudad de España.



¿Por qué?



Respuesta 3: En los números 1 y 3 de la calle Sierpes en Sevilla, España... *La Campana, ¡una excelente bombonería!*



¿Dónde hay una mayor cantidad de información?

•

•

•

Capítulo 6: Teoría de la Información


Página 188

Incertidumbre e información

Ante varios mensajes posibles, en principio todos equiprobables, aquel que tenga una menor probabilidad de aparición será el que contenga una mayor cantidad de información.

- En el ejemplo anterior:
 - Al ser más extenso el número de calles y sus números en una ciudad que el número de ciudades en España, y esto último mayor que los países en Europa, la última respuesta tendrá una mayor incertidumbre.
 - Si suponemos todos los estados equiprobables, entonces la cantidad de información de la respuesta tercera será mayor que las demás.

Las siguientes diapositivas resumen el estudio de Claude Shannon sobre la entropía en su artículo “A Mathematical Theory of Communication” que puede descargarlo en formato pdf desde esta dirección:

<http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html> 

© Jorge Ramío Aguirre

Madrid (España) 2006

•

•

•

Capítulo 6: Teoría de la Información

Página 189

Concepto de variable aleatoria

- Sea X una variable aleatoria con n estados posibles con $X = x_i$ una ocurrencia i ésima:

$$X = \{x_1, x_2, x_3, \dots, x_{n-1}, x_n\}$$

$$p_1 = p(x_1), p_2 = p(x_2), \dots, p_n = p(x_n)$$

Como:

$$0 \leq p_i \leq 1 \quad \text{para } i = 1, 2, \dots, n$$

Entonces:

$$\sum_{i=1}^n p_i = 1$$

La probabilidad de que ocurra p_1 o p_2 o p_3 , etc. será siempre la unidad porque seguro será uno de ellos.


© Jorge Ramío Aguirre

Madrid (España) 2006

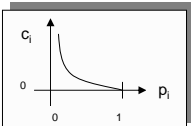
Capítulo 6: Teoría de la Información Página 190

Definición de cantidad de información

- Definiremos c_i a la cantidad de información del estado i , como el logaritmo en base dos de la probabilidad de que ocurra el estado i ésimo.



$c_i = -\log_2(p_i)$



- Logaritmo: $p(x_i) = 1 \Rightarrow$ no hay incertidumbre: $c_i = 0$
 $p(x_i) = 0 \Rightarrow$ máxima incertidumbre: $c_i \rightarrow \infty$
- Signo: $p(x_i) < 1 \Rightarrow \log p(x_i)$ será negativo
- Base 2: Un fenómeno binario \Rightarrow dos estados (bit)

© Jorge Ramío Aguirre Madrid (España) 2006

Capítulo 6: Teoría de la Información Página 191

Grado de indeterminación

$$c_i = \frac{\text{Grado de indeterminación previo}}{\text{Grado de indeterminación posterior}}$$

En una bolsa hay dos papeles con círculos, dos con cuadrados y dos con triángulos: negros o blancos.
 Sacamos a ciegas tres papeles cualesquiera...
 Sea ésta será la combinación elegida...

Combinación 1 ○ □ △

Combinación 2 ○ □ ▲

Combinación 3 ○ ■ △

Combinación 4 ○ ■ ▲

Combinación 5 ● □ △

Combinación 6 ● □ ▲

Combinación 7 ● ■ △

Combinación 8 ● ■ ▲

Si hay equiprobabilidad entonces $p(x_i) = 1/8$

¿Qué cantidad de información tiene cada uno de los estados?

© Jorge Ramío Aguirre Madrid (España) 2006

Capítulo 6: Teoría de la Información Página 192

La incertidumbre del ejemplo del mago

<p>Combinación 1 ○ □ △</p> <p>Combinación 2 ○ □ ▲</p> <p>Combinación 3 ○ ■ △ ✓</p> <p>Combinación 4 ○ ■ ▲</p>	<p>Combinación 5 ● □ △</p> <p>Combinación 6 ● □ ▲</p> <p>Combinación 7 ● ■ △</p> <p>Combinación 8 ● ■ ▲</p>
---	---

Como $p(x_i) = 1/8$ entonces
 Incertidumbre inicial $I_i = 8$
 Daremos algunas pistas 🗯:

- Las figuras no son del mismo color: I_i baja de 8 a 6 al descartarse las combinaciones 1 y 8.
- El círculo es blanco: I_i baja de 6 a 3 (descartamos 5, 6 y 7).
- Hay dos figuras blancas: I_i baja de 3 a 2 (descartamos 4).
- El cuadrado es negro: I_i baja de 2 a 1 (descartamos 2.)

Se acaba la incertidumbre pues la solución es la combinación 3.

Veamos esto ahora matemáticamente ...

→

© Jorge Ramío Aguirre Madrid (España) 2006

Capítulo 6: Teoría de la Información Página 193

Solución matemática al ejemplo del mago

- Las figuras no son del mismo color. I_i baja de 8 a 6:
 $c_{i1} = \log(8/6) = \log 8 - \log 6$
- El círculo es blanco. I_i baja de 6 a 3:
 $c_{i2} = \log(6/3) = \log 6 - \log 3$
- Hay dos figuras blancas. I_i baja de 3 a 2:
 $c_{i3} = \log(3/2) = \log 3 - \log 2$
- El cuadrado es negro. I_i baja de 2 a 1:
 $c_{i4} = \log(2/1) = \log 2 - \log 1$

Todas las magnitudes se pueden sumar como escalares:

$c_i = c_{i1} + c_{i2} + c_{i3} + c_{i4} = \log 8 - \log 1 = \log 8$

© Jorge Ramío Aguirre Madrid (España) 2006

Capítulo 6: Teoría de la Información Página 194

Base del logaritmo

Sean I_i la indeterminación inicial
 I_f la indeterminación final

$$c_i = \log (I_i / I_f) = \log I_i - \log I_f$$

La cantidad de información tiene como unidad de medida la de un fenómeno de sólo dos estados, un fenómeno binario. Luego:

$$c_i = \log_b (2/1) = \log_b 2 - \log_b 1$$

- Si $\log_b 2$ debe ser igual a 1 entonces la base $b = 2$.
- Precisamente a esta unidad se le llama bit (binary digit)
- Ejemplo anterior: $c_i = \log_2 8 = 3$. Es decir, pasamos de la incertidumbre total a la certeza con sólo 3 preguntas.

© Jorge Ramío Aguirre Madrid (España) 2006

Capítulo 6: Teoría de la Información Página 195

Con sólo tres preguntas inteligentes...

Combinación 1 ○ □ △	Combinación 5 ● □ △
Combinación 2 ○ □ ▲	Combinación 6 ● □ ▲
Combinación 3 ○ ■ △ ←	Combinación 7 ● ■ △
Combinación 4 ○ ■ ▲	Combinación 8 ● ■ ▲

Con sólo tres preguntas “*más o menos inteligentes*” podemos pasar de la incertidumbre total a la certeza:

Pregunta 1: ¿Está entre la opción 1 y la 4? ⇒ Sí

Pregunta 2: ¿Está entre la opción 1 y la 2? ⇒ No

Pregunta 3: ¿Es la opción 4? ⇒ No ¡Se acaba la indeterminación!

© Jorge Ramío Aguirre Madrid (España) 2006

Entropía de los mensajes

- Si un fenómeno tiene un grado de indeterminación k y sus estados son equiprobables, la probabilidad p de que se dé uno de esos estados será $1/k$. Luego:

$$c_i = \log_2 (k/1) = \log_2 [1/(1/k)] = -\log_2 p$$

- Si ahora cada uno de estos estados tiene una probabilidad distinta p_i , la entropía H será igual a la suma ponderada de la cantidad de información:

$$H = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_k \log_2 p_k$$

$$H = - \sum_{i=1}^k p_i \log_2 p_i$$

Nota: aunque la ecuación parece bastante lógica, no es inmediata.

http://en.wikipedia.org/wiki/Information_entropy



Definición de entropía

- La entropía de un mensaje X , que se representa por $H(X)$, es el *valor medio ponderado* de la cantidad de información de los diversos estados del mensaje.

$$H(X) = - \sum_{i=1}^k p(x_i) \log_2 p(x_i)$$

Esto lo veremos más adelante...

- Es una medida de la *incertidumbre media* acerca de una variable aleatoria y el *número de bits de información*.



Después del ejemplo de los papeles, podríamos aceptar el concepto de incertidumbre en H . Lo que ahora nos llama la atención 🖐 es lo del número de bits de información.



Propiedades de la entropía

- a) La entropía es no negativa y se anula si y sólo si un estado de la variable es igual a 1 y el resto 0. Esta demostración es sencilla.
- b) La entropía será máxima, hay mayor incertidumbre del mensaje, cuando exista una equiprobabilidad en todos los valores de la variable X. La demostración empírica es muy fácil; no obstante la demostración matemática de este máximo no es directa. El valor máximo de $H(X)$ para una variable de n estados será $\log_2 n$. Si hay n estados equiprobables, entonces $p_i = 1/n$.

Luego:

$$H(X) = - \sum_i p_i \log_2 p_i = - n(1/n) \log_2 (1/n) = - (\cancel{\log_2 1} - \log_2 n)$$

$$H(X)_{\text{máx}} = \log_2 n$$

Concepto codificador óptimo

Nos falta encontrar el segundo término pendiente en la definición de cantidad de información: codificador óptimo.

Introduciendo el signo negativo dentro del logaritmo en la expresión de la entropía, ésta nos quedará como:

$$H(X) = \sum_i p(x) \log_2 [1/p(x)]$$

Veamos un ejemplo de codificación →

La expresión $\log_2 [1/p(x)]$ representará el número necesario de bits para codificar el mensaje X en un codificador óptimo.



Codificador óptimo es aquel que para codificar un mensaje X usa el menor número posible de bits.

Capítulo 6: Teoría de la Información Página 200

Codificación con el método de Huffman

Letra	Frecuencia	Ocurrencias →	Mensaje: MI MAMA ME MIMA
E	1 vez		
I	2 veces		
A	3 veces		
" "	3 veces		
M	6 veces		

Creación del árbol de frecuencias observadas

Código óptimo:

M = 1 " " = 01 A = 000 I = 0010 E = 0011

Mensaje: 1 0010 01 1 000 1 000 01 1 0011 01 1 0010 1 000 (33 bits)

Pregunta: ¿Cuántos bits necesitaría para codificarlo usando ahora código ASCII?

<http://articulos.conclase.net/compresion/huffman.html>

© Jorge Ramío Aguirre Madrid (España) 2006

Capítulo 6: Teoría de la Información Página 201

El número necesario de bits y la entropía

Para que dé un valor exacto, vamos a calcular el número de bits óptimo de codificación para el mensaje M = LELA ELLA (*) de 8 caracteres :

Solución:

$p(L) = 0,5$; $p(E) = 0,25$; $p(A) = 0,25$; y obviamente $\sum p(L, E, A) = 1,0$.

Para codificar L necesitaremos 1 bit: $\log_2 [1/ P(L)] = \log_2 2 = 1$

Para codificar E necesitaremos 2 bits: $\log_2 [1/ P(E)] = \log_2 4 = 2$

Para codificar A necesitaremos 2 bits: $\log_2 [1/ P(A)] = \log_2 4 = 2$

Luego, si L se codifica como 0, E como 10 y A como 11, el mensaje M se codificará como: 0 10 0 11 10 0 0 11, es decir se transmiten 12 bits.

Si calcula la entropía de M obtendrá $H(M) = 1,5$ y al mismo valor se llega con el concepto de número medio de bits: para codificar un mensaje M de 8 elementos, hemos usado 12 bits. Luego $12/8 = 1,5$ bits por elemento.

(*) Mis disculpas este mensaje poco afortunado, pero era difícil encontrar uno con estas características y que tuviese algo de sentido... aunque no sea cierto ☺.

© Jorge Ramío Aguirre Madrid (España) 2006

Entropía condicional: equivocación de X

Si existe una segunda variable Y que influya sobre X, esto nos entregará importante información adicional.

El resultado más interesante es que...

La entropía se reduce: hay más orden y menos incertidumbre.

$$H(X/Y) = - \sum_{x,y} p_{(x,y)} \log_2 p_{(x,y)}$$

Donde $p(x,y) = p(y)p(x/y)$ y la relación $p(x/y)$ es la probabilidad de que se obtenga un estado X conocido el valor de Y.

Luego:

$$H(X/Y) = - \sum_y p_{(y)} \sum_x p_{(x/y)} \log_2 p_{(x/y)}$$

Ejemplo de entropía condicional

Sea $X = \{x_1, x_2, x_3, x_4\}$ con $p(x_i) = 0,25$

Sea ahora $Y = \{y_1, y_2, y_3\}$ con $p(y_1) = 0,5; p(y_2) = 0,25; p(y_3) = 0,25$

Luego $H(X) = 4 \log_2 4 = 2,0$ y $H(Y) = 2 \log_2 4 + \log_2 2 = 1,5$

Suponga además que hay las siguientes dependencias entre X e Y:

Si $Y = y_1 \Rightarrow X = x_1$ o x_2 o x_3 o x_4 (cualquiera con igual probabilidad)

Si $Y = y_2 \Rightarrow X = x_2$ o x_3 (cualquiera con igual probabilidad)

Si $Y = y_3 \Rightarrow X = x_3$ o x_4 (cualquiera con igual probabilidad)

Como $H(X/Y) = - \sum_{y=1}^3 p_{(y)} \sum_{x=1}^4 p_{(x/y)} \log_2 p_{(x/y)}$

$$H(X/Y) = - p(y_1)[p(x_1/y_1)\log_2 p(x_1/y_1) + p(x_2/y_1)\log_2 p(x_2/y_1) + p(x_3/y_1)\log_2 p(x_3/y_1) + p(x_4/y_1)\log_2 p(x_4/y_1)] \\ - p(y_2)[p(x_2/y_2)\log_2 p(x_2/y_2) + p(x_3/y_2)\log_2 p(x_3/y_2) + p(x_4/y_2)\log_2 p(x_4/y_2)] \\ - p(y_3)[p(x_3/y_3)\log_2 p(x_3/y_3) + p(x_4/y_3)\log_2 p(x_4/y_3)]$$

Calculando, se obtiene $H(X/Y) = 1,0 + 0,25 + 0,25 = 1,5$. La entropía de X ha bajado en medio bit por el conocimiento de su relación con Y.

Importancia de la entropía condicional

Equivocación de la clave k
¿Cuál es la probabilidad de que a un criptograma C le corresponda una cifra con una clave k ?

$$H(K/C) = - \sum_c p_{(c)} \sum_k p_{(k/c)} \log_2 p_{(k/c)}$$

Servirá como un parámetro para la evaluación de la fortaleza de un criptosistema según equivocación de clave y mensaje.

Equivocación del mensaje M
¿Cuál es la probabilidad de que a un criptograma C le corresponda un mensaje en claro M ?

$$H(M/C) = - \sum_c p_{(c)} \sum_m p_{(m/c)} \log_2 p_{(m/c)}$$

La ratio r del lenguaje

- Ratio r

- Es el número de “bits de información” en cada carácter para mensajes con una longitud igual a N caracteres. Luego, según la definición de entropía, se tiene:

$$r = H(X)/N \quad (\text{bits/letra})$$

- Si codificáramos un mensaje letra a letra suponiendo además equiprobabilidad entre las letras, se obtiene la denominada ratio absoluta del lenguaje, R :

$$R = H(X) \quad \text{castellano} = 27 \text{ letras}$$

$$R_{\text{castellano}} = \log_2 n = \log_2 27 = 4,75 \quad (\text{bits/letra})$$



•

•

•

Capítulo 6: Teoría de la Información

Página 206

Ratio verdadera del lenguaje

- Ratio verdadera
 - Como las letras que aparecen en un texto no tienen igual probabilidad, su frecuencia de aparición es distinta, los lenguajes están muy estructurados, hay bloques de dos palabras (digramas) característicos, trigramas, poligramas, etc., la ratio baja mucho...

$$1,2 < r < 1,5$$
 - A este valor se llega codificando los mensajes con monogramas, digramas, trigramas, etc., según el estudio hecho por Shannon.

© Jorge Ramío Aguirre

Madrid (España) 2006

• • • • •

•

•

•

Capítulo 6: Teoría de la Información

Página 207

Significado de la ratio del lenguaje

¿Qué significa esto?

- Si un alfabeto consta de L elementos existirán $2^{R \cdot N}$ mensajes posibles de longitud N , la entropía máxima será $H(X)_{\text{máx}} = \log_2 L$, y sólo habrá $2^{r \cdot N}$ mensajes que tengan sentido.

Muy importante: No significa que podamos codificar todos los mensajes de 27 caracteres con 2 bits (esto sería imposible ☺). Sólo significa que la información que contiene cada letra es tan sólo de 1,5 bits.

Veamos un ejemplo

© Jorge Ramío Aguirre

Madrid (España) 2006

• • • • •

Ejemplo de la ratio del lenguaje

Un subalfabeto del castellano módulo 27 consta de 5 caracteres: A, E, O, S, y T, todos ellos equiprobables. Podemos aceptarlo como representativo del lenguaje; es más o menos cierto. De acuerdo, estoy jugando con algo de trampa pero es para que el ejemplo entre justo en una diapositiva ☺.

Pregunta: ¿Cuántos mensaje de longitud 4 existen y cuántos con sentido?

Solución:

$R = \log_2 5 = 2,3219$. Existirán así $2^{R*4} = 2^{2,3219*4} = 625 = 5^4$ mensajes.

Como $1,2 < r < 1,5$ entonces cabe esperar x mensajes con sentido de longitud 4 del orden: $2^{1,2*4} < x < 2^{1,5*4}$ es decir $27 < x < 64$.

Buscando en un diccionario (puede hacerlo) encontramos las 45 palabras que se indican, y que casualmente es el valor medio $(27 + 64)/2 = 45$:

aeta, asas, asea, asee, aseó, ases, asta, atea, atas, ates, ateo, atoa, atoe, atoo, osas, oses, osos, oste, otea, otee, oteo, easo, esas, eses, esos, esta, este esto, etas, tasa, tase, taso, teas, tesa, tese, teso, teta, seas, seso, seta, seto, sosa, sota, sote, soto.

Redundancia del lenguaje

- La redundancia D del lenguaje será la diferencia entre la ratio absoluta y la ratio real:

$$D = R - r$$

$$3,25 < D < 3,55$$

¿Qué significa esto?

- El número de bits extras (*bits redundantes*) necesarios para codificar un mensaje suponiendo un alfabeto de 27 caracteres (codificación con 5 bits puesto que $2^5 = 32$ y $2^4 = 16$) será aproximadamente igual a 3,5.
- D/R será un factor proporcional, luego:

$$68,42 < \% \text{ Red. Lenguaje } (D/R) < 74,73$$

¿No le resulta familiar este porcentaje de reducción en los archivos zip?

http://es.wikipedia.org/wiki/Compresi%C3%B3n_de_datos



¿Es nuestro lenguaje redundante?

- El estudio de Shannon demuestra que es la estructura del lenguaje la que produce esta redundancia:
 - Existe diferencias en la frecuencia de aparición de cada una de las letras de un texto, entregando una distribución típica, como puede ver en las tablas del capítulo 21 de este libro.
 - Existe gran cantidad de digramas comunes (en, es, ...), también muchos trigramas (ado, ida, ...), tetragramas (ando, lado, ...), algunos pentagramas (mente, ...), etc.
 - Existe una estructuración típica de frases y oraciones con sentido en nuestro lenguaje.

Esto dará pistas al criptoanalista para atacar un sistema. Y nuestra misión es crear algoritmos que sean seguros y eviten estos ataques.

Un ejemplo de redundancia (parte 1)

Todos los lenguajes serán redundantes. Esto quiere decir que la misma cantidad de información se puede entregar con menos símbolos o bits.

Sea el siguiente mensaje $M = \text{HBNVZNCRC}$

1ª ayuda:

“En el mensaje original se han quitado las vocales”.

Esto nos permite suponer que entre consonantes habrá 0, 1, 2, 3 y hasta 4 vocales, según las reglas del lenguaje...

$M = _H_B_N_V_Z_N_C_R_C_$



Un ejemplo de redundancia (parte 2)

Teníamos el mensaje $M = \text{HBNVZNCRC}$ y además:

$M = _H_B_N_V_Z_N_C_R_C_$

2ª ayuda:

“El mensaje original contiene cinco palabras”.

Esto nos permite limitar el número de mensajes posibles que tengan sentido. En estas condiciones podrían existir muchos mensajes de 5 palabras, aunque no cumpliesen de forma lógica con las reglas del lenguaje. Un ejemplo válido pero sin sentido lógico podría ser...

$M = \underline{\text{AHÍ}} \underline{\text{BUENO}} \underline{\text{AVE}} \underline{\text{ZONA}} \underline{\text{CERCA}}$



Un ejemplo de redundancia (parte 3)

Teníamos el mensaje $M = \text{HBNVZNCRC}$ y además

$M = _H_B_N_V_Z_N_C_R_C_$

$M = \underline{\text{AHÍ}} \underline{\text{BUENO}} \underline{\text{AVE}} \underline{\text{ZONA}} \underline{\text{CERCA}}$



3ª ayuda y siguientes:

- a) “El mensaje original tiene que ver con un circo”.
- b) “Corresponde al estribillo de una canción infantil”.
- c) “Los espacios están en: $M = \text{HB N VZ N CRC}$ ”.

Seguro que habrá adivinado ya el mensaje.... ☺

$M = \text{HABÍA UNA VEZ UN CIRCO}$



Redundancia y entropía condicional

El ejemplo anterior, además de demostrar que todos los lenguajes son redundantes, es un claro exponente de lo que se entiende en la práctica por entropía condicional.

Cada vez que vamos dando nuevas pistas, disminuye la incertidumbre del mensaje hasta que ésta se anula y por lo tanto la entropía es igual a 0 ya que existe un único mensaje posible con probabilidad igual a la unidad.

Algo parecido ocurre cuando resolvemos un crucigrama y lo anteriormente resuelto nos sirve como pistas para descubrir palabras nuevas. Mientras más palabras tengamos, más fácil se hace avanzar en su resolución. En algunos casos, cuando se ataca una cifra, el criptoanalista usará métodos similares.

Secreto de un sistema criptográfico

Shannon midió el secreto de un criptosistema como la incertidumbre del mensaje en claro conocido el criptograma recibido:

Mensajes $M = \{M_1, M_2, \dots, M_3\}$ $\sum_M p(M) = 1$

Criptogramas $C = \{C_1, C_2, \dots, C_3\}$ $\sum_C p(C) = 1$

Claves $K = \{K_1, K_2, \dots, K_3\}$ $\sum_K p(K) = 1$

¿Cuándo tendrá nuestro sistema un secreto perfecto?




Definiciones previas secreto criptográfico

- $p(M)$: Probabilidad de enviar un mensaje M . Si hay n mensajes M_i equiprobables, $p(M_i) = 1/n$.
- $p(C)$: Probabilidad de recibir un criptograma C . Si cada uno de los n criptogramas recibidos C_i es equiprobable, $p(C_i) = 1/n$.
- $p_M(C)$: Probabilidad de que, a partir de un texto en claro M_i , se obtenga un criptograma C_i .
- $p_C(M)$: Probabilidad de que, una vez recibido un criptograma C_i , éste provenga de un texto claro M_i .

Secreto criptográfico perfecto (1)

Un sistema tiene secreto perfecto si el conocimiento del texto cifrado no nos proporciona ninguna información acerca del mensaje. Es decir, cuando la probabilidad de acierto al recibir el elemento $i + 1$ es la misma que en el estado i .



$$\text{Secreto perfecto} \Rightarrow p(M) = p_C(M)$$

La probabilidad p de enviar un mensaje M con texto en claro $p(M)$ o probabilidad a priori será igual a la probabilidad p de que, conocido un criptograma C , éste se corresponda a un mensaje M cifrado con la clave K . Esta última o probabilidad a posteriori es $p_C(M)$.

Secreto criptográfico perfecto (2)

La probabilidad p de recibir un texto cifrado C al cifrar un mensaje M usando una clave K será $p_M(C)$. Luego, M debe haberse cifrado con alguna clave K :

$$p_M(C) = \sum_K p(K) \quad \text{donde } E_K(M) = C$$

$$\exists k_j / E_{k_j}(M_i) = C_i$$

En el fondo esto viene a significar que para lograr un secreto perfecto, el espacio de claves debe ser al menos de igual tamaño que el espacio de mensajes.

Secreto criptográfico perfecto (3)

La condición necesaria y suficiente del secreto perfecto es que para cualquier valor de M se cumpla que la probabilidad de recibir C , resultado de la cifra de un mensaje M con una clave K , sea la misma que recibir el criptograma C , resultado de la cifra de otro mensaje M' distinto, cifrado con otra clave.

$$p_M(C) = p(C)$$

para todo valor de M

→ Veamos algunos ejemplos

Capítulo 6: Teoría de la Información Página 220

Cifrado con secreto perfecto

Sea el siguiente escenario:

$$p(M) = 1/3 \quad \text{para todo } M \quad p(C) = 1/3$$

$$p_M(C) = 1/3 \quad p_C(M) = 1/3$$

© Jorge Ramió Aguirre Madrid (España) 2006

Capítulo 6: Teoría de la Información Página 221

Cifrado sin secreto perfecto (1)

Sea ahora el siguiente escenario:

$p(M_1) = 1/3$

$p(M_2) = 1/3$

$p(M_3) = 1/3$

Algo más

$p(C_1) = 3/9$

$p(C_2) = 2/9$

$p(C_3) = 2/9$

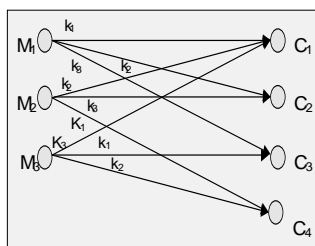
$p(C_4) = 2/9$

¿Probabilidad de que un mensaje M_i se convierta en un criptograma C_i : $[P_{M_i}(C_i)]$ y que un criptograma C_i sea el resultado de la cifra de un mensaje M_i : $[P_{C_i}(M_i)]$?

© Jorge Ramió Aguirre Madrid (España) 2006

Cifrado sin secreto perfecto (2)

Esquema anterior:



$p_{C1}(M_1) = 1/3$	$p_{C1}(M_2) = 1/3$	$p_{C1}(M_3) = 1/3$
$p_{C2}(M_1) = 1/2$	$p_{C2}(M_2) = 1/2$	$p_{C2}(M_3) = 0$
$p_{C3}(M_1) = 1/2$	$p_{C3}(M_2) = 0$	$p_{C3}(M_3) = 1/2$
$p_{C4}(M_1) = 0$	$p_{C4}(M_2) = 1/2$	$p_{C4}(M_3) = 1/2$



$p_{M1}(C_1) = 1/3$	$p_{M1}(C_2) = 1/3$	$p_{M1}(C_3) = 1/3$	$p_{M1}(C_4) = 0$
$p_{M2}(C_1) = 1/3$	$p_{M2}(C_2) = 1/3$	$p_{M2}(C_3) = 0$	$p_{M2}(C_4) = 1/3$
$p_{M3}(C_1) = 1/3$	$p_{M3}(C_2) = 0$	$p_{M3}(C_3) = 1/3$	$p_{M3}(C_4) = 1/3$

La distancia de unicidad

- Se entenderá por Distancia de Unicidad al bloque N de texto cifrado o criptograma mínimo necesario para que se pueda intentar con ciertas expectativas de éxito un ataque en búsqueda de la clave usada para cifrar.
- Este valor se obtiene cuando la equivocación de la clave $H_C(K)$ se acerca a cero o tiende a anularse.
- A medida que se tenga un criptograma más largo, y por tanto más información, se supone que la tarea de ataque del criptoanalista se va facilitando.
- Se busca el tamaño N de criptograma que permita esperar que la solución de K sea única. Suponiendo un cifrador aleatorio, llegamos al modelo de la diapositiva siguiente.

<http://www.cs.ucla.edu/~jkong/research/security/shannon1949/node14.html>



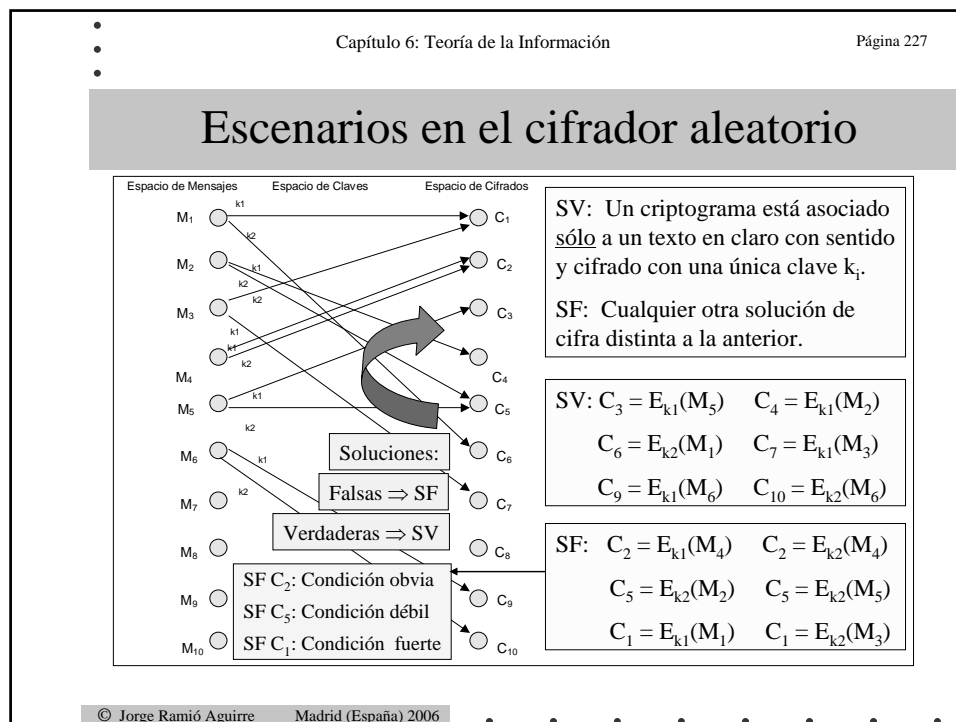
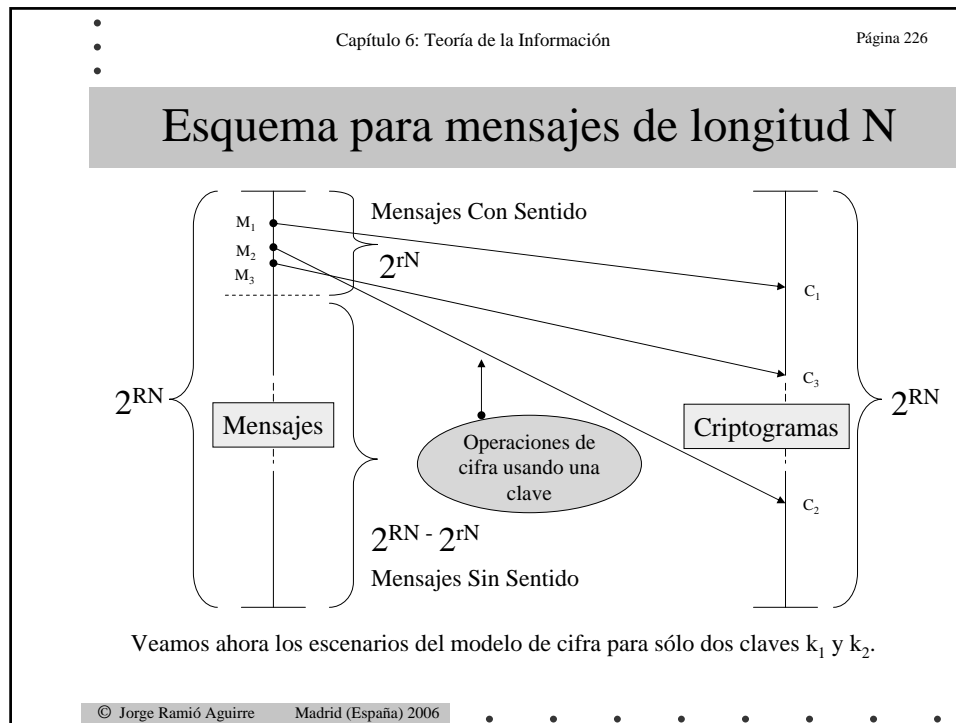
Parámetros del modelo aleatorio (1)

- Existirán 2^{rN} mensajes posibles de longitud N .
- Existirán 2^{rN} mensajes de longitud N con sentido.
- El espacio de mensajes de longitud N se dividirá en:
 - Espacio de los mensajes con sentido: $M_{CS} = 2^{rN}$.
 - Espacio de los mensajes sin sentido: $M_{SS} = 2^{rN} - 2^{rN}$.
- Los 2^{rN} mensajes con sentido serán equiprobables siendo su valor $p(M_{CS}) = 1/2^{rN} = 2^{-rN}$.
- El resto de mensajes ($2^{rN} - 2^{rN}$) correspondientes a aquellos sin sentido tendrán una probabilidad nula $p(M_{SS}) = 0$, ya que nunca serán generados.

Parámetros del modelo aleatorio (2)

- Existirán $2^{H(K)}$ claves equiprobables.
- En donde $H(K)$ es la entropía de la clave.
- Con $p(K) = 1/2^{H(K)} = 2^{-H(K)}$.
- Con estas claves se cifrarán todos los mensajes con sentido dando lugar a 2^{rN} textos cifrados posibles de longitud N .
- A diferencia de los mensajes, como es lógico los criptogramas obtenidos serán todos equiprobables.

En la siguiente diapositiva se muestra un esquema de este modelo.



Cálculo de la distancia de unicidad (1)

- Para cada solución correcta de un texto M cifrado con una clave k del espacio $2^{H(K)}$, existirán otras $(2^{H(K)}-1)$ claves con la misma probabilidad de entregar una solución falsa SF.

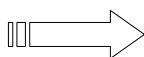
Sea q la probabilidad de obtener un mensaje con sentido:

$$q = 2^{rN} / 2^{RN} = 2^{(r-R)N} = 2^{-DN}$$

Luego:

$$SF = (2^{H(K)}-1) q = (2^{H(K)}-1) 2^{-DN} = 2^{H(K) - DN} - 2^{-DN}$$

$$SF \approx 2^{H(K) - DN}$$



$$\log_2 SF = H(K) - DN$$

Cálculo de la distancia de unicidad (2)

La solución $SF = 0$ es imposible porque sólo se llega a ella de forma asintótica con un valor de N infinito como se muestra en la diapositiva siguiente.

Se acepta entonces que haya como máximo una sola solución falsa, de ahí su nombre de unicidad, luego:

$$SF = 2^{H(K) - DN} \quad \text{Si hacemos } SF = 1 \Rightarrow H(K) - DN = 0$$

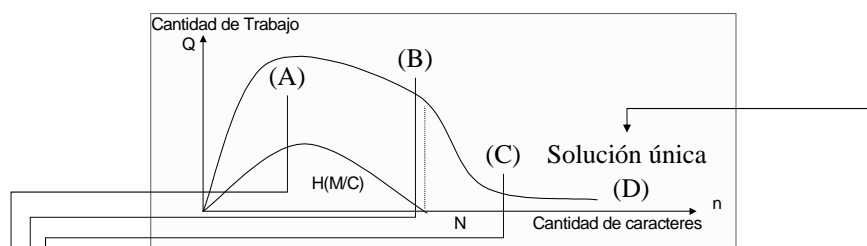
Por lo tanto:

$$N = H(K) / D$$



El valor N será el número mínimo de bytes o caracteres que deberá tener el criptograma C para intentar un ataque por estadísticas del lenguaje. Por lo general el valor real necesario de N será unas 10 veces superior.

Cantidad de trabajo Q en un criptoanálisis



- (A) Inicialmente hay que hacer un arduo trabajo para obtener algo coherente. Nos encontraremos con muchas soluciones falsas.
- (B) Cuando se tiene una cantidad “adecuada” de texto cifrado, la cantidad de trabajo disminuye. Se descartan algunas soluciones.
- (C) Cuando se anula la equivocación de la clave, $H(M/C) = 0$, disminuyen las soluciones falsas y la solución tiende a ser única.

Algunos ejemplos de distancia de unicidad

- Para el cifrador del César módulo 27 en el que “la clave” es b, todos los posibles desplazamientos de caracteres, $1 \leq b \leq 26$, su entropía $H(X) = \log_2 26 = 4,7$ bits por lo que $N = 4,7/3,4 = 1,4$ caracteres.
- Para el mismo cifrador del César pero con clave, si el alfabeto tiene n caracteres, existirán n! claves posibles. En este caso la entropía de la clave puede aproximarse como $H(X) = \log_2 27! \approx 27 * \log_2 (27/e)$, por lo que $N = 27 * \log_2 (27/2,72)/3,4 = 27,4$ caracteres.
- En el sistema DES la clave verdadera es de 56 bits por lo que su entropía $H(X) = 56$. Si el mensaje sólo contiene letras mayúsculas (27 elementos) podríamos decir que $N = 56/3,4 = 16,5$ caracteres.
- Nota: aunque el valor de N sea ahora más bajo no quiere decir en absoluto que el DES sea menos seguro que el cifrador del César con clave. Este último se puede atacar fácilmente con estadísticas del lenguaje muy elementales y el DES no. Además, recuerde que se debe contar con un criptograma varias veces mayor que el valor de N si desea que su criptoanálisis tenga alguna posibilidad de éxito.

El uso de técnicas de difusión

Para lograr un mayor secreto en las operaciones de cifra, Shannon propuso usar dos técnicas: difusión y confusión.

Difusión: es la transformación sobre el texto en claro con el objeto de dispersar las propiedades estadísticas del lenguaje sobre todo el criptograma. Se logra con transposiciones.

TRANSPOSICIONES

La transposición consiste básicamente en una permutación, es decir, cambiar los caracteres de lugar según una regla, una función, etc. Por ejemplo el carácter primero se posiciona en el lugar cuarto, el segundo en el lugar tercero, etc.

El uso de técnicas de confusión

Confusión: transformación sobre el texto en claro con objeto de mezclar los elementos de éste, aumentando la complejidad de la dependencia funcional entre la clave y el criptograma. Se obtiene a través de sustituciones.

SUSTITUCIONES

La sustitución consiste básicamente modificar la información, es decir, sustituir un carácter por otro de acuerdo a una regla, una función, etc. Por ejemplo cambiar la letra A por la letra M, la letra B por la letra X, etc.

Ambas técnicas se usan en sistemas clásicos orientados a caracteres y también en los modernos pero en este caso operando sobre bits.

Fin del capítulo

Cuestiones y ejercicios (1 de 2)

1. Al despertar ponemos la radio y escuchamos noticias que no nos llaman la atención. ¿Por qué decimos que no había información?
2. Justifique la definición logarítmica de cantidad de información, es decir la razón de que $c_i = -\log(p_i)$.
3. ¿Por qué usamos la base 2 en el logaritmo que define c_i ?
4. ¿Cuál es el número mínimo -e inteligente- de preguntas que hay que hacer para pasar de la incertidumbre a la certeza en un sistema de n estados equiprobables? ¿Y si ahora no son equiprobables?
5. ¿Por qué la entropía es no nula y se anula si y sólo si uno de los estados de la variable es igual a la unidad?
6. Codificamos en binario un sistema con 256 estados equiprobables. Si no usamos un codificador óptimo, ¿cuántos bits son necesarios? Mediante un codificador óptimo, ¿usaremos más o menos bits?

Cuestiones y ejercicios (2 de 2)

7. ¿Qué representa la expresión $\log_2 [1/p(x)]$ en la entropía $H(X)$? Si $p(x_1)=0,6$; $p(x_2)=0,3$; $p(x_3)=0,1$ calcule $\log_2 [1/p(x)]$. ¿Qué opina?
8. Definimos un alfabeto con 71 elementos (mayúsculas y minúsculas, minúsculas acentuadas, dígitos, punto, coma). Si estos elementos son equiprobables, ¿cuál es la ratio absoluta de este alfabeto?
9. ¿La ratio verdadera es mayor o menor que la absoluta? ¿Por qué?
10. Un alfabeto consta de 8 elementos equiprobables. ¿Cuántos posibles mensajes de tamaño 4 existen? De éstos, ¿cuántos mensajes podrían tener sentido si esos 8 elementos representan al idioma castellano?
11. ¿Cuándo decimos que un sistema tiene secreto perfecto? En un sistema real, ¿es eso posible? Piense en algún ejemplo y coméntelo.
12. ¿Por qué se dice que hay que minimizar las soluciones falsas SF en el modelo aleatorio para romper la clave? ¿Es la clave k única?


•
•
•

Capítulo 6: Teoría de la Información

Página 236

Use el portapapeles

Prácticas del tema 6 (1/1)

Software CripClas: http://www.criptored.upm.es/software/sw_m001c.htm 

1. Encuentre la entropía del mensaje $M = \text{MI MAMA ME MIMA}$, compárela con el resultado de la diapositiva correspondiente, 33 bits para codificar 15 caracteres: $33/15 = 2,2$. ¿Por qué no coinciden? Repita este cálculo ahora con el mensaje $M = \text{RARORARO}$ y saque conclusiones.
2. Encuentre la entropía de $M = \text{ABCDEFGHIJKLMNÑOPQRSTUVWXYZ}$ es decir el alfabeto en castellano módulo 27, y compárela con el valor que aparece en la dispositiva correspondiente.
3. ¿Cómo son las entropías de $M = \text{TE AMO}$ y $M = \text{Te amo}$? ¿Por qué?
4. Copie en el portapapeles todas estas preguntas, guarde el archivo con el nombre `prtema6.txt` y encuentre su entropía. Encuentre luego la entropía de otros archivos `txt`, grandes y pequeños, y saque conclusiones.
5. Encuentre la frecuencia de monogramas del archivo anterior, `prtema6.txt`. Compárela en la misma pantalla con la tabla de frecuencias estándar.

© Jorge Ramío Aguirre Madrid (España) 2006