Conceptos relacionados

CURSO CRIPTOGRAFÍA Y SEGURIDAD INFORMÁTICA

Ana I. González-Tablas Ferreres

José María de Fuentes García-Romero de Tejada

Lorena González Manzano

Pablo Martín González

ÍNDICE

- 3. Conceptos relacionados
  - Teoría de la información
    - Entropía
    - Entropía condicionada
  - Aleatoriedad
  - Complejidad algorítmica

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ÍNDICE

- 3. Conceptos relacionados
  - **Teoría de la información**
    - Entropía
    - Entropía condicionada
  - Aleatoriedad
  - Complejidad algorítmica

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

TEORÍA DE LA INFORMACIÓN

- Bases matemáticas (Claude E. Shannon)
  - *A mathematical theory of communication*, Bell Syst. Tech. J., vol.23. 1948

- Fundamentos teóricos de la criptografía: Criptología científica

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

TEORÍA DE LA INFORMACIÓN

- Establece una métrica para evaluar el secreto de un cifrador
- Se basa en la incertidumbre que sobre el texto en claro tiene un criptoanalista que intercepta un texto cifrado

  – Cifrador incondicionalmente seguro
    - No se filtra nada, independientemente de la longitud de C (Vernam)

  – Cifrador matemáticamente vulnerable
    - Cuanta mayor sea la longitud de C, mayor cantidad de información se filtra (y por tanto está disponible para el criptoanalista)

COSEC uc3m
Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

CANTIDAD DE INFORMACIÓN

- Sea $M=\{m_1, m_2, \ldots ,m_n\}$ una fuente de mensajes estadísticamente independientes cuyas probabilidades de ocurrencia respectivas son:

  $p(m_1),\ldots, p(m_n)$ con $\Sigma\, p(m_i)=1$

- La cantidad de información (c_i) de un mensaje m_i es:

  $$\boxed{c_i= - \log_2 p(m_i) \text{ bits}}$$

- A mayor $p(m_i)$, menor c_i

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ÍNDICE

- 3. Conceptos relacionados
  - **Teoría de la información**
    - **Entropía**
    - Entropía condicionada
  - Aleatoriedad
  - Complejidad algorítmica

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ENTROPÍA

- Entropía de una fuente M es la cantidad promedio de información transportada por un mensaje perteneciente a dicha fuente

- Entropía de la fuente M:

$$\boxed{H(M) = - \Sigma\, p(m_i) \log_2 p(m_i) \ \text{bits}}$$

- Bit: entropía de una fuente con 2 mensajes equiprobables

  $- (1/2 \log_2 ½ + 1/2 \log_2 ½) = 1/2 \log_2 2 + 1/2 \log_2 2) =1$ bit

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ENTROPÍA

- Entropía es la cantidad de información que es previsible ganar tras la aparición de un m_i
- La entropía de M mide la incertidumbre que, a priori, tiene un observador acerca de la aparición de un m_i
- A mayor entropía, mayor incertidumbre sobre M

Entropía cero = incertidumbre cero = $p(m_i)=1$ para algún *i*

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ENTROPÍA

- Sea $M=\{m_1, m_2, \ldots, m_n\}$ con $\Sigma\, p(m_i)=1$

- Propiedades
  1. $0 \leq H(M) \leq \log_2 n$
  2. $H(M)= 0$ si y sólo si $p(m_i)=1$ para algún i
  3. $H(M)= \log_2 n$ si y sólo si $p(m_i)=1/n$ para $1 \leq i \leq n$

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ENTROPÍA

- Ej. Considere una fuente con 2 elementos $M=\{m_1, m_2\}$ con $p(m_1)=1/3$ y $p(m_2)=2/3$. Calcule la entropía de M

$H(M) = - \Sigma\ p(m_i) \log_2 p(m_i) = 1/3 \log_2 3 - 2/3 \log_2 2/3 =0.52 + 0.38 = 0.9$

- Ej. Considere una fuente con 2 elementos $M=\{m_1, m_2\}$ con $p(m_1)=0.4$ y $p(m_2)=0.6$. Calcule la entropía de M

$H(M) = - \Sigma\ p(m_i) \log_2 p(m_i) = -0.4 \log_2 0.4 – 0.6 \log_2 0.6 =0.52 + 0.44 = 0.96$

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ÍNDICE

- 3. Conceptos relacionados
  - – **Teoría de la información**
    - Entropía
    - **Entropía condicionada**
  - – Aleatoriedad
  - – Complejidad algorítmica

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ENTROPÍA CONDICIONADA

- Cuando existe alguna relación entre las apariciones de dos mensajes consecutivos n_j (de una fuente N) y m_i (de una fuente M), la presencia del primero disminuye la incertidumbre del segundo

- La entropía de M condicionada por N, **H(M|N),** se define como el valor medio de la cantidad de información de M conocido N

$$\boxed{H(M|N) = - \sum_{j} p(n_j) \sum_{i} p(m_i|n_j) \log_2 p(m_i|n_j)}$$

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ENTROPÍA CONDICIONADA

- Ej. $M=\{m_1, m_2, m_3, m_4\}$, $p(m_1)=p(m_2)=p(m_3)=p(m_4)=1/4$ y $N=\{n_1, n_2\}$, $p(n_1)=p(n_2)=1/2$.

  $N=n_1 \Rightarrow M=m_1$ ó m_2 (equiprobablemente)

  $N=n_2 \Rightarrow M=m_3$ ó m_4 (equiprobablemente)

- $H(M)=2$ y
- $H(M|N)= 1/2(1/2 \lg_2 2+1/2 \lg_2 2)+ 1/2(1/2 \lg_2 2+1/2 \lg_2 2)=1$

- El conocimiento de N hace disminuir la entropía resultante de M

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ENTROPÍA CONDICIONADA

- Los métodos criptográficos tratan de maximizar H(M|N) siendo M el conjunto de textos en claro y N el de los cifrados

- Todos los cifradores (menos Vernan) filtran alguna información sobre el texto en claro al texto cifrado, y según la longitud del texto cifrado crece, mayor es la información filtrada

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ÍNDICE

- 3. Conceptos relacionados
  - Teoría de la información
    - Entropía
    - Entropía condicionada
  - **Aleatoriedad**
  - Complejidad algorítmica

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

VARIABLE ALEATORIA

- Sea S un espacio muestral con distribución de probabilidad P (cada posible valor que X puede tomar en S tiene asociada una determinada probabilidad)
- Una variable aleatoria X es una función de S al conjunto de los números reales X : $S \rightarrow E = \mathbb{R}$
- Ejemplo de variable aleatoria discreta
- S = {cara, cruz} $\qquad X(s) = \begin{cases} 1 & \text{si } s=\text{cara} \\ 0 & \text{si } s=\text{cruz} \end{cases}$
-
- Si moneda equilibrada: P(X=1) = ½ ; P(X=0) = ½

COSEC uc3m

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

SECUENCIA ALEATORIA

- Múltiples usos
  - Distribución de claves
  - Protocolos de autenticación mutua
  - Generación de claves de sesión
  - Generación de claves para RSA
  - Generación de flujos de bits para algoritmos de cifrado simétrico de flujo
- Criterios de aleatoriedad:
  - Distribución uniforme : La frecuencia de aparición de unos y ceros debe ser aproximadamente la misma
  - Independencia: Ninguna subsecuencia puede ser inferida de otras

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

SECUENCIA ALEATORIA

- Baterías de tests
  - Existen test para probar distribución uniforme

- No existen test para probar independencia
- Existen test para demostrar la no independencia

- **Si no pasa tests, aleatoriedad descartada**
- **Si pasa todos los tests, no se puede garantizar aleatoriedad**
  - El Maurer Universal Test no es definitivo

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ALEATORIEDAD
SECUENCIA ALEATORIA

- Las aplicaciones criptográficas generalmente utilizan algoritmos para generar números “aleatorios”
- Aunque una secuencia verdaderamente aleatoria no puede estar generada por un algoritmo dado que éste por definición es determinista
- Diferencia:
  - Pseudoaleatoriedad (PRNG)
    - Algoritmo
  - Aleatoriedad (TRNG) [Uso de fuentes no deterministas]
    - Fuente de entropía tomada de ciertos procesos naturales
    - Eliminación del sesgo con funciones resumen

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

ÍNDICE

- 3. Conceptos relacionados
  - Teoría de la información
    - Entropía
    - Entropía condicionada
  - Aleatoriedad
  - **Complejidad algorítmica**

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

COMPLEJIDAD ALGORÍTMICA

- Campo de la matemática que estudia los algoritmos bajo la dificultad de su resolución

- Clasifica los algoritmos según su complejidad

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

PROBLEMAS Y ALGORITMOS

- Problema
  - Planteamiento de una tarea en un determinado contexto

- Algoritmo
  - Conjunto finito de <u>operaciones</u>, que realizadas en un determinado <u>orden</u>, resuelven un <u>problema</u>
  - Los algoritmos pueden trabajar sobre un ejemplo particular de problema (<u>problemas particulares</u>)
  - Si un algoritmo resuelve todos los problemas particulares

    ⇒ el algoritmo resuelve el <u>problema genérico</u>

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

PROBLEMAS Y ALGORITMOS

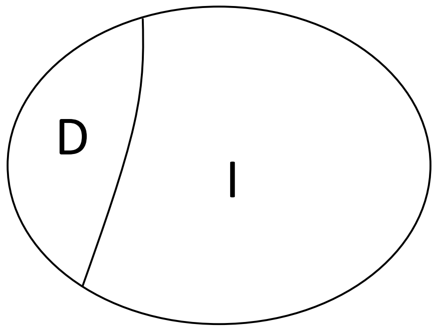
- Turing demuestra que no todos los problemas tienen un algoritmo que los resuelva

- ¡No todos los problemas tienen solución!

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

PROBLEMAS Y ALGORITMOS

- Una primera clasificación de los problemas
  - **Indecidibles (I)**
    - No resolubles mediante un algoritmo
  - **Decidibles (D)**
    - Cuentan con al menos un algoritmo para su resolución



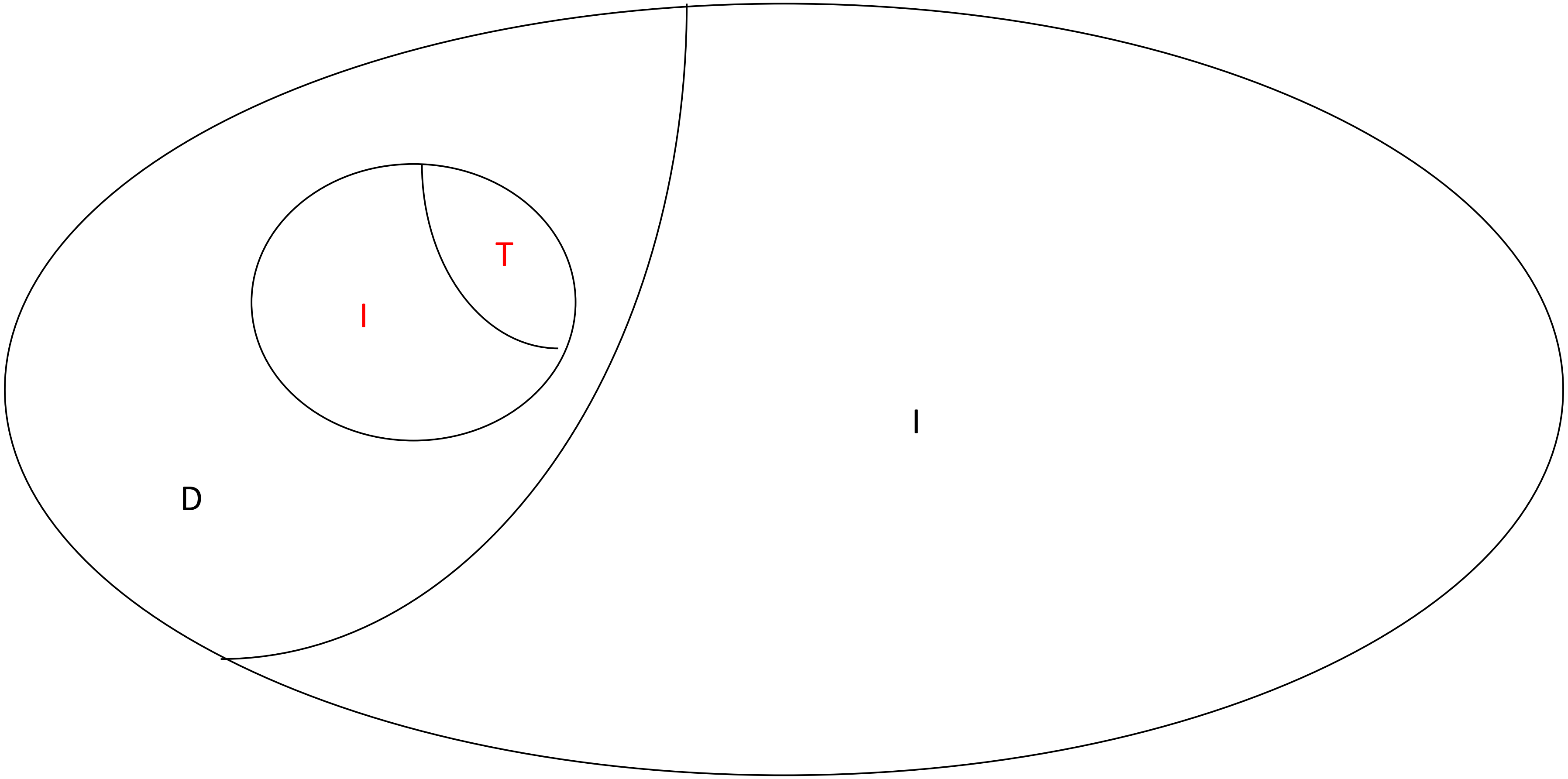
Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

PROBLEMAS TRATABLES E INTRATABLES

- Existen problemas cuya solución es inabordable por el elevado número de operaciones a realizar

- Una segunda clasificación de los problemas:
  - Intratables (I)
    - No es factible obtener su solución en un tiempo razonable con potencia de cálculo actual
  - Tratables (T)
    - Existen al menos un algoritmo que resuelve cualquier problema particular en tiempo razonable

PROBLEMAS TRATABLES E INTRATABLES



TIEMPO DE EJECUCIÓN

- La dificultad para resolver una instancia de un problema se mide según su tiempo de ejecución (t)
- Es función del tamaño de la entrada (n)
- Se analiza el comportamiento del algoritmo cuando n crece (comportamiento asintótico)
  - Se dice que un algoritmo presenta una complejidad polinómica si el tiempo t es de orden polinómico o menor
    - Logarítmico O(log n): Ej. $t = 5 \log n$ O(log n)
    - Potencia de n (polinómico) $O(n^c)$: Ej. $t = 2n^3 + 6n$ $O(n^3)$
  - vs. complejidad exponencial si el tiempo t es de orden mayor que polinómico
    - Exponencial $O(c^n)$: Ej. $t = 3^n + 4n$ $O(3^n)$
    - Factorial $O(n!)$: Ej. $t = 5n! + 6^n$ $O(n!)$

TIEMPO DE EJECUCIÓN

- En un ordenador con 1 millón de operaciones por segundo

Tamaño n	$\log_2 n$ (t)	n (t)	n^2 (t)	2^n (t)
10	$3 \cdot 10^{-6}$ s	10^{-5}s	10^{-4} s	10^{-3}s
10^2	$7 \cdot 10^{-6}$ s	10^{-4}s	10^{-2} s	10^{14}siglos
10^3	$10 \cdot 10^{-6}$ s	10^{-3}s	1 s	Muy grande
10^4	$13 \cdot 10^{-6}$ s	10^{-2}s	1,7 min	Muy grande
10^5	$17 \cdot 10^{-6}$ s	10^{-1}s	2,8 h	Muy grande

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

CLASES DE COMPLEJIDAD ALGORÍTMICA

- Un problema puede resolverse por distintos algoritmos
- Los problemas se clasifican en clases de complejidad según el tiempo en el que pueden ser resueltos:
  - Clase P (Polynomial time)
  - Clase NP (Non deterministic Polynomial time)
  - Otras clases…

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

CLASES DE COMPLEJIDAD ALGORÍTMICA

- Problemas de Clase P (Polynomial time)
  - Son problemas <u>Tratables</u>
  - Se resuelven mediante <u>algoritmos polinómicos</u> (buenos algoritmos)
  - Los algoritmos utilizados son <u>deterministas</u>
    - En cada paso de computación se determina de forma única el siguiente paso
  - La concatenación de dos algoritmos P es otro algoritmo P

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

CLASES DE COMPLEJIDAD ALGORÍTMICA

- Problemas de clase NP (Non deterministic Polynomial time)
  - Contiene problemas Intratables (y tratables)
    - ¿P $\subset$ NP?
  - Los problemas intratables se resuelven mediante algoritmos no polinomiales (malos algoritmos), como los exponenciales
  - Los algoritmos utilizados son no deterministas
    - En cada paso de computación necesitan una selección entre diferentes opciones
  - Ejemplos:
    - Problema del logaritmo discreto → Diffie-Hellman, ElGamal
    - Problema de la factorización → RSA

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Pablo Martín González

CURSO CRIPTOGRAFÍA Y SEGURIDAD INFORMÁTICA

COSEC

uc3m | Universidad Carlos III de Madrid