Related concepts

CRYPTOGRAPHY AND COMPUTER SECURITY

Ana I. González-Tablas Ferreres
José María de Fuentes García-Romero de Tejada
Lorena González Manzano
Sergio Pastrana Portillo

uc3m | Universidad Carlos III de Madrid

COSEC

OUTLINE

- 3. Related concepts
  - Information theory
    - Entropy
    - Conditional entropy
  - Randomness
  - Algorithmic Complexity

OUTLINE

- 3. Related concepts
  - Information theory
    - Entropy
    - Conditional entropy
  - Randomness
  - Algorithmic Complexity

INFORMATION THEORY

- Mathematical basis (Claude E.Shannon)
  - *A mathematical theory of communication*, Bell Syst. Tech. J., vol.23.1948

- Theoretical foundations of cryptography: Scientific Cryptology

INFORMATION THEORY

- It is a branch of applied mathematics involving the quantification of information
- It is based on the uncertainty that a cryptanalyst, who has intercepted a piece of ciphertext, has about plaintext
- A cipher is mathematically vulnerable:
  - If increasing C length we obtain additional information
  - It does not necessarily imply that it will be vulnerable in real life → Computational security
- Cipher unconditionally secure (perfect secrecy):
  - Independent of the length of C (Vernam)

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

AMOUNT OF INFORMATION

- Let $M=\{m_1, m_2, \ldots, m_n\}$ be a source of statistically independent messages, which respective occurrence probabilities are:

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

- The amount of information (c_i) in a message m_i is:

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

- When $p(m_i)$ increases, c_i decreases

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

OUTLINE

- 3. Related concepts
  - Information theory
    - Entropy
    - Conditional entropy
  - Randomness
  - Algorithmic Complexity

ENTROPY

- Entropy of a source M is the average amount of information a m_i belonging to that source conveys

- Entropy of source M:

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

- Bit: entropy of a source having 2 equiprobable messages

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

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

ENTROPY

- Entropy of source M is the expected amount of information provided by an observation of M

- The entropy of a source M measures the (a priori) uncertainty about the outcome before an observation of M

- For higher values of entropy, we expect higher values of uncertainty about the source M

Zero entropy = zero uncertainty = $p(m_i)=1$ for some *i*

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

ENTROPY

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

- Properties
  1. $0 \leq H(M) \leq \log_2 n$
  2. $H(M)= 0$ if and only if $p(m_i)=1$ for some i
  3. $H(M)= \log_2 n$ if and only if $p(m_i)=1/n$ for $1 \leq i \leq n$

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

ENTROPY

- Ex. Consider a source with 2 elements M ={m_1, m_2} with $p(m_1)=1/3$ and $p(m_2)=2/3$. Calculate the entropy of M

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

- Ex. Consider a source with 2 elements M ={m_1, m_2} with $p(m_1)=0.4$ and $p(m_2)=0.6$. Calculate the entropy of 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, Sergio Pastrana Portillo

OUTLINE

- 3. Related concepts
  - Information theory
    - Entropy
    - Conditional entropy
  - Randomness
  - Algorithmic Complexity

CONDITIONAL ENTROPY

- If there exists a bound between the appearances of two consecutive messages n_j (belonging to source N) and m_i (belonging to source M), the presence of the first message lowers the uncertainty for the second one.

- We define entropy of M conditional on N, **H(M|N),** as the mean value of the amount of information of M, being N known

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

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

CONDITIONAL ENTROPY

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

  $N=n_1 \Longrightarrow M=m_1$ or m_2 (equiprobably)

  $N=n_2 \Longrightarrow M=m_3$ or m_4 (equiprobably)

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

- As we have seen, if we have information about N, this will result in a reduction of the entropy of M

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

CONDITIONAL ENTROPY

- Cryptographic methods intend to maximize H(M|N), being M the set of plaintexts and N the set of ciphertexts

- Every cipher (with the exception of Vernam) (unintentionally) 'leaks' some plaintext information to ciphertext

- The amount of information 'leaked' increases along with the ciphertext length

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

OUTLINE

- 3. Related concepts
  - Information theory
    - Entropy
    - Conditional entropy
  - Randomness
  - Algorithmic Complexity

RANDOM VARIABLE

- Let S be a sample space following some probability distribution P

- A random variable (stochastic) X is a function from S to the set of real numbers: X : S E=|R
- Example of random discrete variable
- S = {heads, tails} X(s) = 1 si s=head; 0 si s=tail
- If balanced coin: P(X=1) = ½ ; P(X=0) = ½

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

RANDOM SEQUENCE

- Multiple applications:
  - Key distribution
  - Mutual authentication protocols
  - Session key generation
  - RSA key generation
  - Bitstream generation for stream ciphers
- Randomness criteria:
  - Uniform distribution: Frequency of 1's and 0's must be aprox. the same
  - Independence: None of the subsequences can be inferred from others

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

RANDOM SEQUENCE

- Test batteries
  - There are some tests available to prove uniform distribution
    - There are no specific tests to prove independence
    - There exist some tests to prove no independence
    - **If a sequence does not pass the tests, we can state that randomness is discarded**
    - **If all tests are passed, there is no guarantee about randomness**
      - Maurer Universal Test is not definitive

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

RANDOMNESS
RANDOM SEQUENCE

- Cryptographic algorithms do usually use algorithms to generate “random” numbers…
- Although a truly random sequence can never be generated by an algorithm (it is deterministic by definition)
- Difference:
  - Pseudorandomness (PRNG)
    - Algorithm
  - Randomness (TRNG) [Use of nondeterministic sources]
    - Sources of entropy based on unpredictable natural processes
    - Reduction or elimination of bias (eg., with hash functions)

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

OUTLINE

- 3. Related concepts
  - Information theory
    - Entropy
    - Conditional entropy
  - Randomness
  - Algorithmic Complexity

ALGORITHM COMPLEXITY

- Mathematical discipline focused on the analysis of algorithms, dealing with their difficulty of resolution

- It classifies algorithms in terms of their complexity

PROBLEMS AND ALGORITHMS

- Problem
  - A specific task in a determined context

- Algorithm
  - A finite set of operations, which carried out in a determined order,solves a problem
  - An algorithm handles particular cases of the problem (particular problem)
  - In case an algorithm solves all particular problems

    ⇨ the algorithm solves the generic problem

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

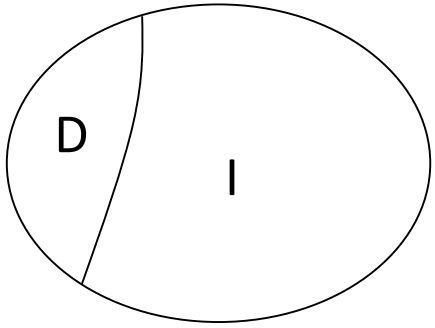
PROBLEMS AND ALGORITHMS

- Turing proved that not every problem has an algorithm capable of solving it

- Not every problem has a solution!

PROBLEMS AND ALGORITHMS

- First classification of problems
  - **Undecidables (U)**
    - Not solvable by means of an algorithm
  - **Decidables (D)**
    - There is, al least, an algorithm that solves them



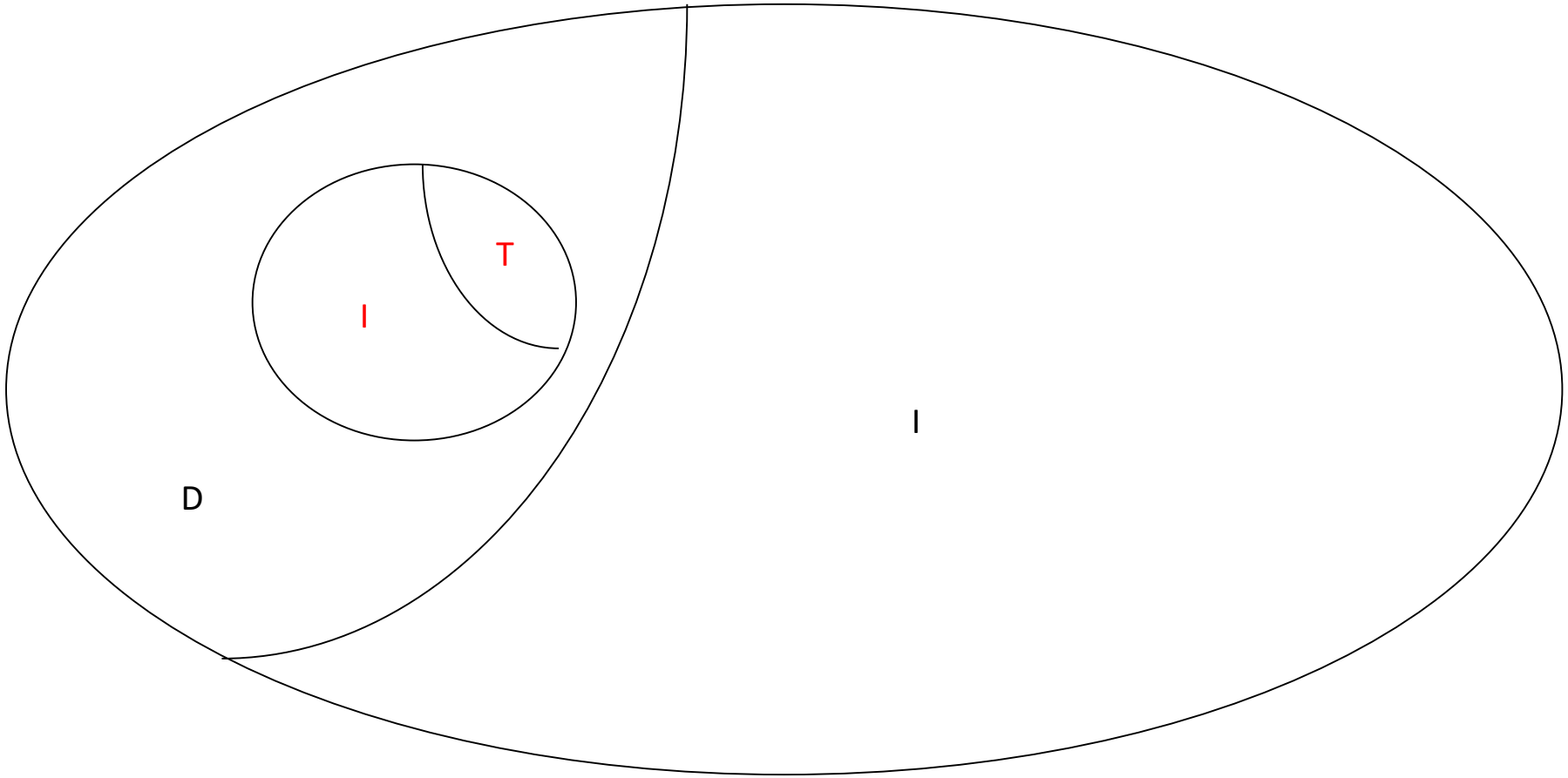
Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

TRACTABLE AND INTRACTABLE PROBLEMS

- There exist unapproachable problems, due to the high number of operations required to solve them

- Second classification of problems:
  - Intractables (I)
    - It is not feasible to reach a solution using state of the art computing power
  - Tractables (T)
    - There exists, at least, an algorithm that solves every particular problem in a reasonable timeframe

TRACTABLE AND INTRACTABLE PROBLEMS



TIME COMPLEXITY

- Time complexity t measures the amount of time taken by an algorithm to solve a problem
- It is a function of the size of the input (n)
- Asymptotic behaviour when n grows
  - Polynomial time algorithm
    - Asymptotic behaviour $O(n^c)$ [or smaller]
    - Eg polynomial order
      - Ex $t = 2n^3 + 6n$ $O(n^3)$ t:execution time
  - Non Polynomial time algorithm
    - Asymptotic behaviour $O(c^n)$ [or greater tan polynomial]
    - Eg exponential order
      - Ex. $t = 3^n + 4n$ $O(3^n)$

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

TIME COMPLEXITY

- Using a computer able to execute 1 million operations per second

Size 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} centuries
10^3	$10 \cdot 10^{-6}$ s	10^{-3}s	1 s	Very big
10^4	$13 \cdot 10^{-6}$ s	10^{-2}s	1,7 min	Very big
10^5	$17 \cdot 10^{-6}$ s	10^{-1}s	2,8 h	Very big

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

ALGORITHMIC COMPLEXITY CLASSES

- A problem can be solved using different algorithms
- Problems are divided into complexity classes:
  - P Class (Polynomial)
  - NP Class (Non deterministic Polynomial problems)
  - Other…

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

ALGORITHMIC COMPLEXITY CLASSES

- P Class (Polynomial)
  - Solve Tractables problems
  - By means of polynomial algorithms (good algorithms)
  - Uses deterministic algorithms
    - Each step is determined in a deterministic manner
  - The result of concatenating two P algorithms is another P algorithm

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

ALGORITHMIC COMPLEXITY CLASSES

- **NP Class (Non deterministic Polynomial problems)**
  - Solve Intractables problems (and tractable)
    - P $\subset$ NP?
  - By means of non polynomial algorithms (like exponential algorithms (bad algorithms))
  - Uses non-deterministic algorithms
    - Need to select different alternatives before reaching a solution
  - Examples:
    - Discrete logarithm problem
    - Factoring problem

Ana I. González-Tablas Ferreres, José M. de Fuentes García-Romero de Tejada, Lorena González Manzano, Sergio Pastrana Portillo

CRYPTOGRAPHY AND COMPUTER SECURITY

COSEC

uc3m | Universidad Carlos III de Madrid