PhD course: Finite Fields
Some slides for 4th Lecture

Diego Ruano

Department of Mathematical Sciences
Aalborg University
Denmark

18-12-2012

Polynomials can be considered as functions

Actually, any function over a finite field can be represented by a polynomial, thanks to Lagrange interpolation

The Lagrange interpolant

$$l_i = \prod_{0 \leq j < n, j \neq i} \frac{x - u_j}{u_i - u_j}$$

has the property that $l_i(u_j) = 0$ if $i \neq j$ and $l_i(u_i) = 1$.

For arbitrary $v_0, \ldots v_{n-1}$, the Lagrange polynomial

$$f = \sum_{0 \leq i < n} v_i l_i$$

verifies $f(u_i) = v_i$ for all i.

Complexity

Evaluating a polynomial $f \in F[X]$ of degree less than n at n distinct points $u_0, \ldots, u_{n-1}$ takes $2n^2 - 2n$ operations and the Lagrange interpolation takes $7n^2 - 8n + 1$ operations.

One can also understand evaluation as the F-linear map:

$$(f_0, \ldots, f_{n-1}) \mapsto \Big(\sum_{0 \le j < n} f_j u_0^j, \ldots \sum_{0 \le j < n} f_j u_{n-1}^j \Big)$$

that can be represented using the Vandermonde matrix

$$\begin{pmatrix} 1 & u_0 & u_0^2 & \cdots & u_0^{n-1} \\ 1 & u_1 & u_1^2 & \cdots & u_1^{n-1} \\ 1 & u_2 & u_2^2 & \cdots & u_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_{n-1} & u_{n-1}^2 & \cdots & u_{n-1}^{n-1} \end{pmatrix}$$

and interpolation by the inverse matrix.

Lagrange interpolation is unique

2 proofs:

- Consider f_1, f_2 two interpolation polynomials with degree lower than n that interpolate $n+1$ points. Then $f_1 - f_2$ has $n+1$ roots, contradiction.
- Consider the Vandermonde Matrix, since it is non-singular the system of equations has unique solution.

Secret sharing schemes

Example: Dankort PIN number for 4 friends.

Let t, w be positive integers $t \leq w$. A (t, w)-threshold scheme is a method of sharing a key K among a set of w participants (denoted by $\mathcal{P}$), in such a way that any t participants can compute the value of K but no group of $t-1$ participants can do so.

The value of K is chosen by a special participant D called the Dealer. We assume $D \notin \mathcal{P}$. He secretly gives the participants some information called share.

Trivial secret sharing: $t = w$

We consider $t = w$. We need all participants to recover the secret.

- Encode the secret as an integer s.
- Give to each player i (except one) a random integer x_i.
- Give to the last player the number $s - x_1 - x_2 - \ldots - x_{n-1}$.
- The secret is the sum of the players' shares.

Binary version:

- Encode the secret as an arbitrary length binary number s.
- Give to each player i (except one) a random number x_i with the same length as s.
- Give to the last player $s \text{ XOR } x_1 \text{ XOR } x_2 \text{ XOR } \ldots \text{ XOR } x_{n-1}$
- The secret is the bitwise XOR of all the players' shares.

Shamir (t, w)-Threshold Scheme

Initialization Phase:

1. D chooses w distinct, non-zero elements of $\mathbb{F}_p$, denoted x_i, $1 \leq i \leq w$ (ASSUMPTION: $p \geq w + 1$).
2. For $1 \leq i \leq w$, D gives the value x_i to P_i. The values x_i are public.

Share Distribution:

3. Suppose D wants to share a key $K \in \mathbb{F}_p$. D secretly chooses (independently at random) $t - 1$ elements of $\mathbb{F}_p$, which are denoted by $a_1, \ldots, a_{t-1}$.
4. For $1 \leq i \leq w$, D computes $y_i = f(x_i)$, where

$$f(x) = K + \sum_{j=1}^{t-1} a_j x^j \bmod p$$

5. For $1 \leq i \leq w$, D gives the share y_i to P_i.

Shamir (t, w)-Threshold Scheme

How do t participants recover the secret?

Why cannot $t - 1$ participants recover the secret?

HINT: Lagrange interpolation

Extensions

- We desired that any t of the w participants should be able to determine the key K.
- A more general situation is to specify exactly which subsets of participants should be able to determine the key and which should not.

Example (from Wikipedia):

Imagine that the Board of Directors of a company would like to protect their secret formula. The president of the company should be able to access the formula when needed, but in an emergency any 3 of the 12 board members would be able to unlock the secret formula together. This can be accomplished by a secret sharing scheme with $t = 3$ and $w = 15$, where 3 shares are given to the president, and 1 is given to each board member.

How to do to this in general? Secret Sharing for General Access Structures.

Steganography

- Steganography is derived from Greek words *steganos* and *graphein* and means covered writing.
- Cryptography only protects the information but Steganography protects both the messages and the communicating parties.
- Steganography is the science of communicating secret messages in such a way that no one, apart from the sender and receiver, can detect the existence of a message.
- It is the science of invisible communication, even the fact of communicating has to be kept secret.

Steganography

- The secret message we want to protect is embedded into an apparently innocuous object, the cover.
- Covers can be images, texts, emails, computer files.
- The typical classic example is invisible ink in a handwritten document.
- Ancient Greeks used the trick of shaving the head of slaves and tattoed a messaged on it, after the hair had grown the message was hidden.
- Acrostic: first letter or word of each, line, paragraph,...

Digital Steganography: today's typical cover is a computer file: a document, image, a program or a protocol. Media files are ideal for steganographic purposes because of their large size and redundancy.

Steganography ⟷ Watermarking

A watermark is a hidden message that stores information about the cover in which it is contained and which is the main value to be protected: ownership, copyrights, ...

- Presence of a watermark into a cover is (in general) not a secret.
- The main requirement of watermarking is imperceptibility.

For steganography: the cover has no intrisic value it is just a container to hide the message. And we require that third parties should not be able to determine whether the cover contains a hidden information.

Example: Mimic functions

- A mimic function f changes a file F to acquire the statistical properties of another file F'.
- This means that each string s that occurs in F with probability p also occurs in F' with approximately equal probability.
- This method avoid attacks based on statistical analysis.

Grammar-based mimicking used grammar models to transform a text imitating a default grammatical structure. Example SpamMimic:

http://www.spammimic.com

We make the cover and we speak about cover synthesis.

Hiding information in images

- The sender uses an innocuous digital image which is stored in the computes as a sequence of bits corresponding to its pixels.
- Then the sender adjusts the color of some pixels in such a way to correspond to a letter in the alphabet.
- Usually, the least significant bits of selected pixels are replaced. Hence, this technique is called LSB Steganography.
- The change of the color is almost imperceptible for a human eye and moreover, there is no way to compare it with the previous image.

We slightly modify the cover and we talk about cover modification.

This will be our framework in this lecture.

Steganographic schemes

A steganographic scheme is formed by the embedding and recovering functions. These functions may depend on a key used to increase the system security. Formally it is a five-tuple $\mathcal{S} = (\mathcal{C}, \mathcal{M}, \mathcal{K}, \mathrm{emb}, \mathrm{rec})$, where

- $\mathcal{C}$ is a set of possible covers
- $\mathcal{M}$ is a set of possible messages
- $\mathcal{K}$ is a set of possible keys
- $\mathrm{emb} : \mathcal{C} \times \mathcal{M} \times \mathcal{K} \to \mathcal{C}$ is an embedding function
- $\mathrm{rec} : \mathcal{C} \times \mathcal{K} \to \mathcal{M}$ is a recovering function s.t.

$$\mathrm{rec}(\mathrm{emb}(c, m, k), k) = m, \text{ for all } m \in \mathcal{M}, c \in \mathcal{C}, k \in \mathcal{K}$$

The original cover c is called plain cover and $\mathrm{emb}(c, m, k)$ is called stegocover.
There are many choices for $\mathcal{C}, \mathcal{M}, \mathcal{K}$... we consider:

Steganographic schemes

- We assume that the cover $c \in \mathbb{F}_q^n$ and the secret $m \in \mathbb{F}_q^k$
- A digital steganographic scheme $\mathcal{S}$ of type $[n, k]$ over the alphabet $\mathbb{F}_q$ is a pair of functions
  1. $\mathrm{emb} : \mathbb{F}_q^n \times \mathbb{F}_q^k \to \mathbb{F}_q^n$ is an embedding function
  2. $\mathrm{rec} : \mathbb{F}_q^n \to \mathbb{F}_q^k$ is a recovering function s.t.

$$\mathrm{rec}(\mathrm{emb}(c, m)) = m, \text{ for all } m \in \mathbb{F}_q^k, c \in \mathbb{F}_q^n$$

- The set of possible messages is $\mathbb{F}_q^k$ and the elements of $\mathbb{F}_q^n$ are called cover vectors (or cover sequences).
- The original message m is hidden in the cover vector c as $c' = \mathrm{emb}(c, m)$ and recovered as $\mathrm{rec}(c')$.

Parameters

- Cover length n
- Embedding capacity k
- Embedding radius ρ: maximum number of embedding changes

$$\rho = \max\{d(c, \mathrm{emb}(c, m)) : c \in \mathbb{F}_q^n, m \in \mathbb{F}_q^k\}$$

- The average number of embedding changes R_α. Assuming al messages and covers equally probably

$$R_a = \frac{1}{q^{kn}} \sum_{c \in \mathbb{F}_q^n, m \in \mathbb{F}_q^k} d(c, \mathrm{emb}(c, m))$$

One me say that we have an $[n, k, \rho]$ stegoscheme or a $[n, k, R_a]$ stegoscheme.

Good protocol

A good protocol has to fulfil:

- To have efficient embedding and retrieval algorithms.
- To have good parameters: k/n as big as possible and ρ/n as small as possible.

The protocol is said to be proper if the number of changes produced in the cover by the embedding process is the minimum possible allowed by the covering map, i.e. if

$$d(c, \mathrm{emb}(c, m)) = d(c, \mathrm{rec}^{-1}(m)), \text{ for all } c \in \mathbb{F}_q^n, m \in \mathbb{F}_q^k$$

Exercise: The embedding map of any steganogrphic scheme can be slightly modified to make it proper.

Examples

Consider an LSB protocol applied to a bitmap image where each pixel is represented by h bits, we can consider:

- The cover is the whole image, the protocol is the image, the protocol (for each pixel) is $\mathrm{emb} : \mathbb{F}_2^h \times \mathbb{F}_2 \rightarrow \mathbb{F}_2^h$ defined as $\mathrm{emb}((c_1, \ldots, c_h), m))) = (c_1, \ldots, c_{h-1}, m)$ and $\mathrm{rec} : \mathbb{F}_2^h \rightarrow \mathbb{F}_2$ defined as $\mathrm{rec}(y_1, \ldots, y_h) = y_h$. Then it is a $[1, h, 1]$ protocol.
- The cover is the set of all least significant bits, the protocol (for each pixel)$\mathrm{emb} : \mathbb{F}_2 \times \mathbb{F}_2 \rightarrow \mathbb{F}_2$ defined as $\mathrm{emb}(c, m))) = m$ and $\mathrm{rec} : \mathbb{F}_2 \rightarrow \mathbb{F}_2$ defined as $\mathrm{rec}(y) = y$. Then it is a $[1, 1, 1]$ protocol.

Coding theory

Steganographic schemes and Error correcting codes

Consider that we want to embed k bits of information into a string c of n bits allowing one change at most. What is the minimum possible length n of the cover to perform this purpose?

Answer: There $n+1$ possibilities of changing at most one bit of c and 2^k messages of length k, then $n+1 \geq 2^k$.
F5

F5

F5 hides sequences of k bits into covers of $n = 2^{k-1}$ bits.

- For $a \in \mathbb{Z}$, $0 \leq a \leq 2^k - 1$, denote $[a]_2$ binary expresion of a with k bits
- For $a \in \mathbb{F}_2^k$, denote $[a]_{10}$ integer in decimal form.
- $e(i)$ vector of canonical basis of $\mathbb{F}_2^n$ and $e(0) = 0$.

$$\mathrm{emb}(c, m) = c + e\left(\left[m + \sum_{i=1}^{n} c_i [i]_2\right]_{10}\right)$$

$$\mathrm{rec}(c) = \sum_{i=1}^{n} c_i [i]_2$$

Exercise: Check that $\mathrm{rec}(\mathrm{emb}(c, m)) = m$ for all c, m. Relate F5 with a well-known error correcting code, note that the recovering map is linear. Compute its matrix. What is the code? What can be said about the embedding and recovering process?

From Stegoschemes to codes and from Codes to stegoschemes

Proposition

Let $(\mathrm{emb}, \mathrm{rec})$ be a proper $[n, k]$ stegoscheme. For each $m \in \mathbb{F}_q^k$ we consider the code $\mathcal{C}_m = \{x \in \mathbb{F}_q^n : \mathrm{rec}(x) = m\}$. Then the family $\{C_m : m \in \mathbb{F}_q^k\}$ gives a partition on $\mathbb{F}^n$. Furthermore for all $m \in \mathbb{F}_q^k$ the map $\mathrm{dec}_m : \mathbb{F}_q^n \to \mathcal{C}_m$ defined by $\mathrm{dec}_m(x) = \mathrm{emb}(x, m)$ is a decoding map for the code C_m.

Proposition

Let $\{\mathcal{C}_m : m \in \mathbb{F}_q^k\}$ be a family of codes indexed by $\mathbb{F}_q^k$ and giving a partion of $\mathbb{F}_q^n$. For each $m \in \mathbb{F}_q^n$ let dec_m be a minimum distance decoding map of $\mathcal{C}_m$. Consider the maps $\mathrm{emb} : \mathbb{F}_q^n \times \mathbb{F}_q^k \to \mathbb{F}_q^n$ and $\mathrm{rec} : \mathbb{F}_q^n \to \mathbb{F}_q^k$ defined by $\mathrm{emb}(x, m) = \mathrm{dec}_m(x)$ and $\mathrm{rec}(x) = m$ if $x \in \mathcal{C}_m$. Then $(\mathrm{emb}, \mathrm{rec})$ is a proper $[n, k]$ stegoscheme.

Conclusion

The following objects are equivalent

- A proper $[n, k]$ stegoscheme $(\mathrm{emb}, \mathrm{rec})$ over $\mathbb{F}_q$, and
- A family $\{(\mathcal{C}_m, \mathrm{dec}_m) : m \in \mathbb{F}_q^k\}$ indexed by $\mathbb{F}_q^k$, where for every m: $\mathcal{C}_m \subset \mathbb{F}_q^n$ is a code, dec_m is a minimum distance decoding map for $\mathcal{C}_m$ and the family $\{\mathcal{C}_m : m \in \mathbb{F}_q^k\}$ gives a partition of $\mathbb{F}_q^n$.