



UNIVERSITA' DEGLI STUDI DELLA BASILICATA  
Dipartimento di Matematica, Informatica ed Economia  
Corso di Laurea in Scienze e Tecnologie Informatiche

---

Tesi di Laurea

# STEGANOGRAFIA SU PDF

Relatore: Ing. Domenico Daniele Bloisi

Candidato: Rocchina Romano

26 Marzo 2020

---

Anno Accademico 2018 - 2019



UNIVERSITA' DEGLI STUDI DELLA BASILICATA

Dipartimento di Matematica, Informatica ed Economia

Corso di Laurea in Scienze e Tecnologie Informatiche

---

## OUTLINE

- **LA STEGANOGRAFIA**
  - **LE MOTIVAZIONI**
  - **I CONTRIBUTI**
  - **RISULTATI RAGGIUNTI**
  - **CONCLUSIONI E SVILUPPI FUTURI**
  - **DISTRIBUZIONE DEL CODICE SORGENTE**
-

## LA STEGANOGRAFIA

✓ **ETIMOLOGIA:** Il termine “**steganografia**” deriva dal greco “**stèganos**” e “**gráphein**”, che significano rispettivamente “**coperta**” e “**scrittura**”.

✓ **OBIETTIVO :** nascondere l'**esistenza stessa** di una comunicazione segreta.



MESSAGGIO ESTRATTO:

ciao

### ✓ PROPRIETA' DI UN BUON SISTEMA STEGANOGRAFICO

1. CAPACITA' DI INCORPORAMENTO
2. IMPERCETTIBILITA'
3. SICUREZZA
4. RESISTENZA ALLA MANOMISSIONE

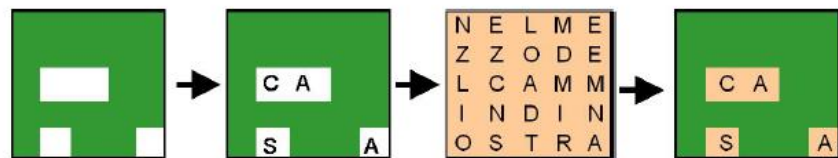
**ATTENZIONE !**

**Non è possibile soddisfare  
contemporaneamente  
tutti e 4 i requisiti!**

## UN PO' DI STORIA

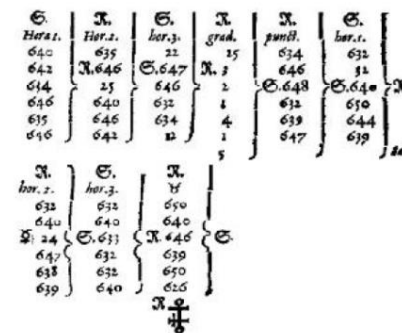
### ➤ Le griglie di Cardano

✓ Si tratta di una **griglia** che veniva appoggiata su un foglio di carta bianca al fine di scrivere il messaggio segreto. Successivamente, la griglia veniva rimossa e si cercava di completare la scrittura del resto del foglio con una frase di senso compiuto.



### ➤ Tritemio: “Steganographia”, 1499

✓ Si tratta di una trilogia, dove i primi 2 libri contengono numerosi metodi per nascondere messaggi segreti all'interno di altre scritture, mentre la terza parte della trilogia sembra un libro di astronomia occulta, ma , in realtà, contiene dei **cifrari di sostituzione numerica**.



### ➤ Le immagini di al-Qaida, 2002

✓ Si presume che abbiano utilizzato la **steganografia per inviare messaggi segreti su siti internet**.

Th	Sch	Tz	Z	X	W	U	T	S	R	Q	P	
01	02	03	04	05	06	07	08	09	10	11	12	
26	27	28	29	30	31	32	33	34	35	36	37	
51	52	53	54	55	56	57	58	59	60	61	62	
76	77	78	79	80	81	82	83	84	85	86	87	
O	N	M	L	I	H	G	F	E	D	C	B	A
13	14	15	16	17	18	19	20	21	22	23	24	25
38	39	40	41	42	43	44	45	46	47	48	49	50
63	64	65	66	67	68	69	70	71	72	73	74	75
88	89	90	91	92	93	94	95	96	97	98	99	00



## ➤ Le cifre nulle, Seconda Guerra Mondiale

- ✓ è una tecnica usata durante la seconda guerra mondiale soprattutto per la trasmissione di messaggi segreti **via radio**.
- ✓ consiste nel nascondere messaggi segreti all'interno di altri messaggi testuali.
- ✓ Per poter risalire al messaggio nascosto, si consideravano, ad esempio, le iniziali oppure la seconda lettera di ogni parola in sequenza.

Ad esempio, il testo, riportato di sotto, è un messaggio che è stato mandato realmente da una spia tedesca:

*“Apparently **ne**utral's **pr**otest **is** **th**oroughly **di**scounted **an**d **ig**nored.  
**Is**man **ha**rd **hi**t. **Bl**ockade **is**sue **af**fects **pr**etext **for** **em**bargo **on** **by** **p**roducts, **ej**ecting  
**su**ets **an**d **ve**getable **oi**ls.”*

Se consideriamo la seconda lettera di ogni parola, otteniamo il seguente messaggio nascosto:

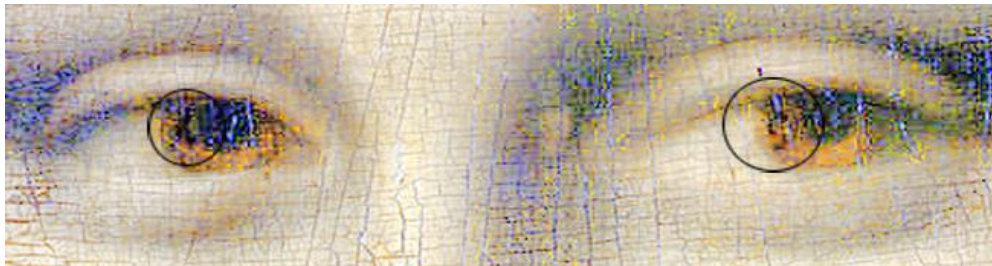
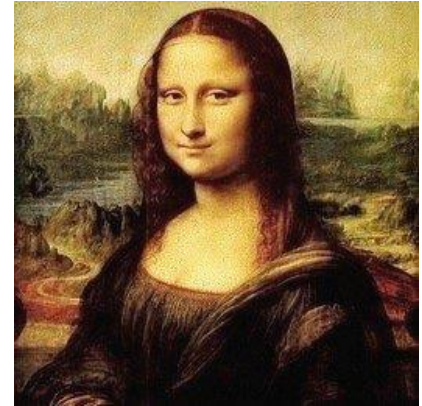
*“**Pershing sails from NY June 1**”*

**NOTA:** in realtà, c'è una "r" di troppo e la "i" alla fine viene interpretata come "1".

## LA STEGANOGRAFIA NELL'ARTE

### ➤ Leonardo Da Vinci: La Gioconda (1503 -1504)

- ✓ Alle spalle della donna , in particolare sotto il **ponte**, sono stati ritrovati **due numeri**, un **7** ed un **2**, che alcuni studiosi hanno ricondotto al 1472, anno in cui il **ponte Gobbo sul Trebbia** venne distrutto dalla piena del fiume.
- ✓ Negli **occhi** della donna sono state ritrovate **due lettere**, una "G" ed una "S", che alcuni studiosi hanno interpretato come le iniziali di **Bianca Giovanna Sforza**, figlia di Ludovico il Moro, signore di Bobbio.







UNIVERSITA' DEGLI STUDI DELLA BASILICATA

Dipartimento di Matematica, Informatica ed Economia

Corso di Laurea in Scienze e Tecnologie Informatiche

---

## LE MOTIVAZIONI



**ATTENZIONE !**

**In alcuni Paesi l'utilizzo della Steganografia è ILLEGALE!!!**

- **Protezione del Copyright**

- **Privacy**

✓ **ESEMPIO:** garantire la **riservatezza** di informazioni strettamente personali, come ad esempio nel caso dei **referti medici**, per i quali si avverte la necessità di mantenere segreta l'identità del paziente.

---

## IL PROBLEMA DEL PRIGIONIERO (Simmons, 1983)

- Alice e Bob sono 2 **prigionieri**, che devono escogitare un piano per poter **fuggire**.
- Essi si scambiano dei messaggi attraverso il **guardiano** Wendy
- Se Wendy scopre che Alice e Bob si stanno scambiando dei **messaggi segreti**, uno dei due verrà messo in isolamento ed il **piano fallirà**
- Per questo motivo, devono trovare un modo per nascondere il messaggio segreto in un testo apparentemente “innocuo”







# UNIVERSITA' DEGLI STUDI DELLA BASILICATA

Dipartimento di Matematica, Informatica ed Economia  
Corso di Laurea in Scienze e Tecnologie Informatiche

---

## I CONTRIBUTI

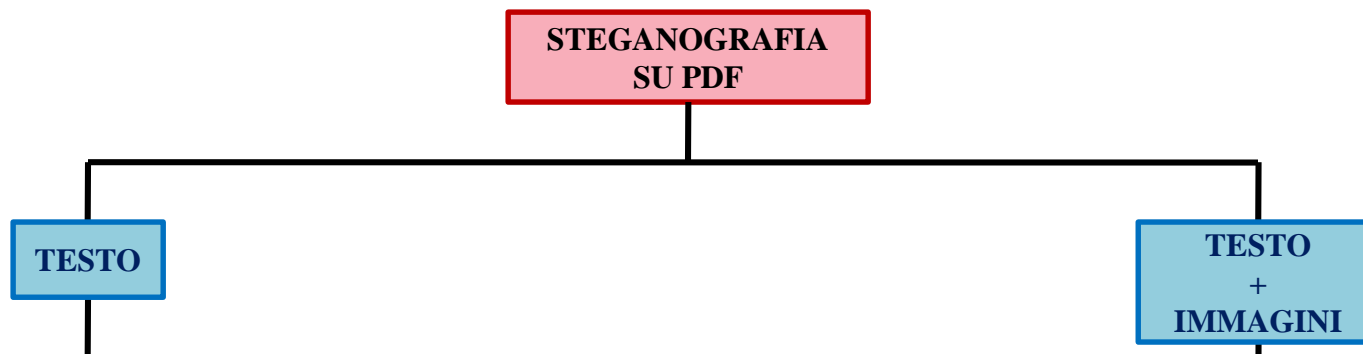
- E' stata sviluppata una nuova tecnica di **STEGANOGRAFIA SU TESTO**
- che permette di **INSERIRE** ed **ESTRARRE** un messaggio segreto
- utilizzando il **formato PDF** (“**Portable Document Format**”)

## STRUMENTI UTILIZZATI:

### ✓ iText:

- È una **libreria** che fornisce una serie di metodi per la **lettura**, la **scrittura** e la **modifica** di file **PDF**;
  - disponibile sia per **Java**, sia per **.NET**
  - Ultima versione disponibile è la **7**
  - Distribuita sotto licenza **AGPL** (“**Affero General Public License** “)
-

## STEGANOGRAFIA SU PDF



### SCHEMA CODIFICA / DECODIFICA

SEQUENZA BINARIA	DIMENSIONE IMMAGINI BIANCHE
00	4x4
01	8x8
10	16x16
11	32x32

### SCHEMA CODIFICA / DECODIFICA

SEQUENZA BINARIA	DIMENSIONE IMMAGINI BIANCHE
00	4x4
01	8x8
10	16x16
11	32x32

ALGORITMO  
F5

## PROCESSO DI EMBEDDING

✓ processo che si occupa di “**incorporare**” il messaggio segreto nel **cover PDF**

### INPUT

#### 1. MESSAGGIO SEGRETO:

CIAO

#### 2. COVER PDF

"Nel mezzo del cammin di nostra vita  
mi ritrovai per una selva oscura,  
chè la diritta via era smarrita."

PROCESSO  
DI  
EMBEDDING

### OUTPUT

#### STEGO PDF

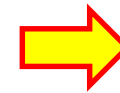
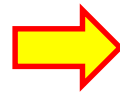
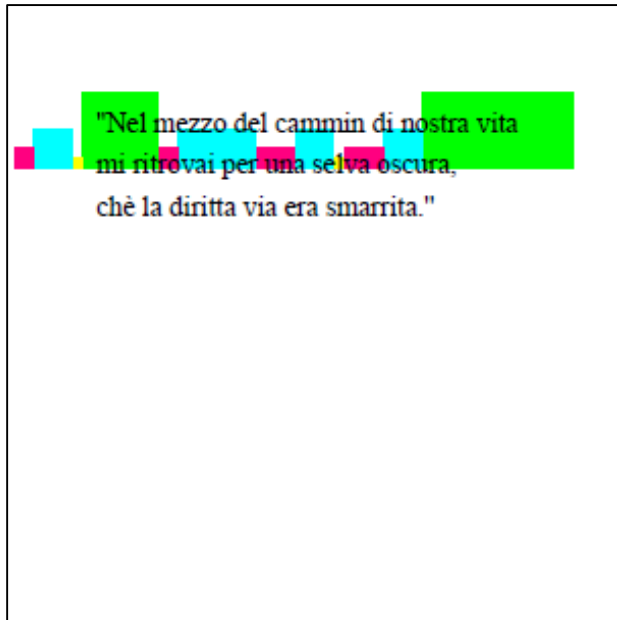
"Nel mezzo del cammin di nostra vita  
mi ritrovai per una selva oscura,  
chè la diritta via era smarrita."

## PROCESSO DI ESTRAZIONE

- ✓ processo che si occupa di “**estrarre**” il messaggio segreto a partire da uno **stego PDF**

**INPUT**

**STEGO PDF**



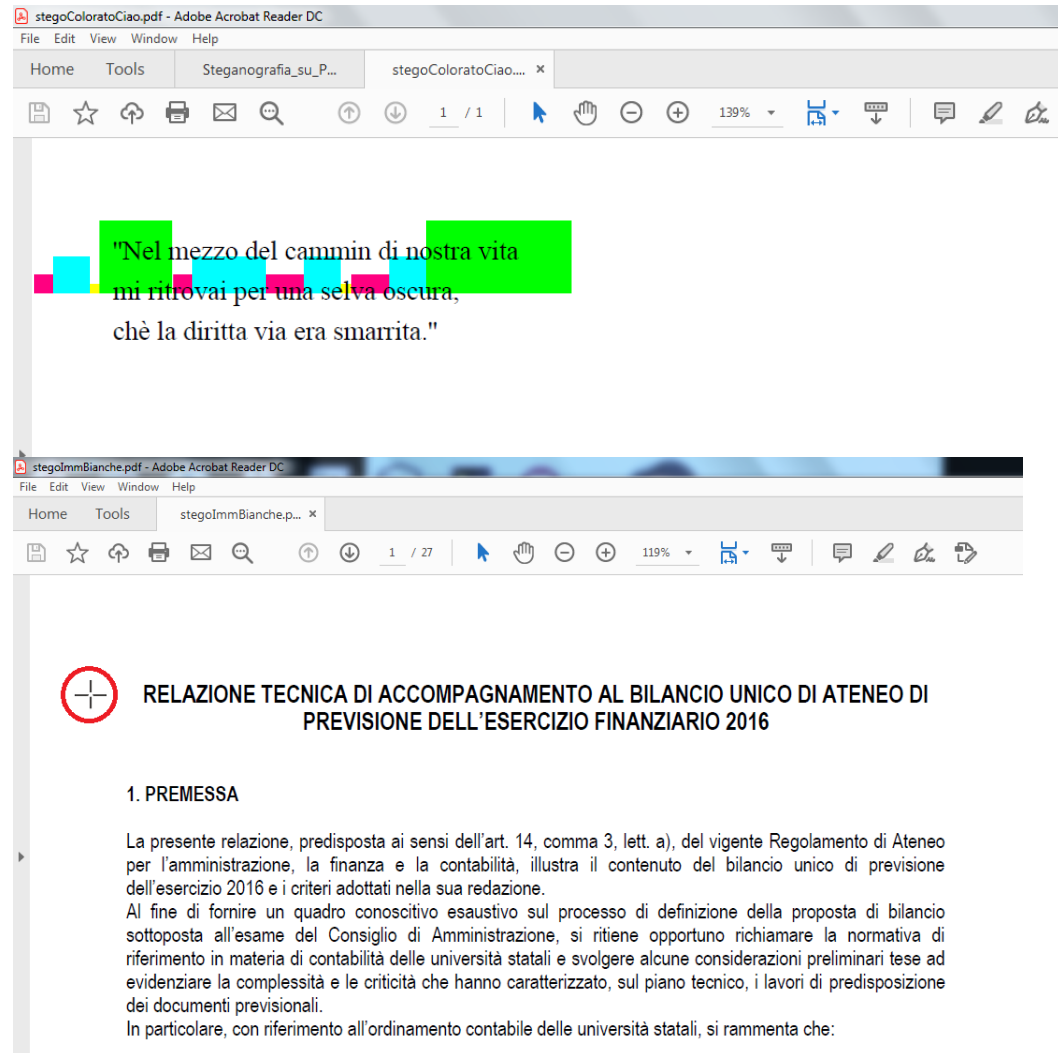
**OUTPUT**

**MESSAGGIO SEGRETO  
ESTRATTO**



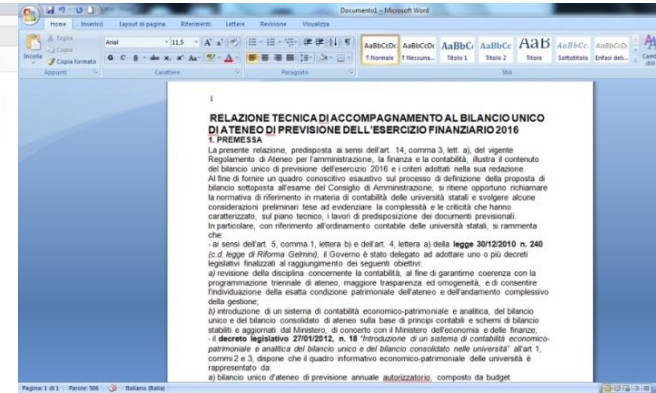
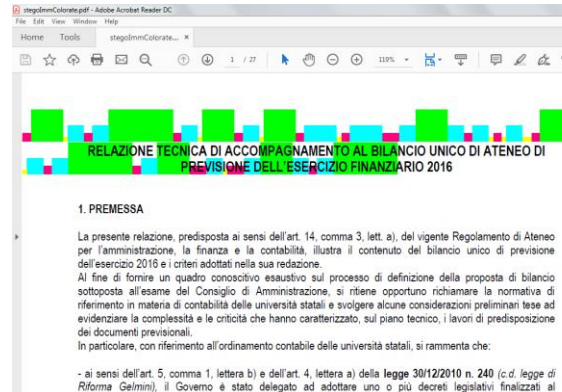
## RISULTATI RAGGIUNTI 1/3

- 1) **iText** permette di inserire le immagini bianche **sottoponendole** al contenuto originale del documento PDF in modo tale da **non oscurarlo**;
- 2) le **immagini bianche**, inserite nello stego PDF, vengono **rilevate dal mouse**;

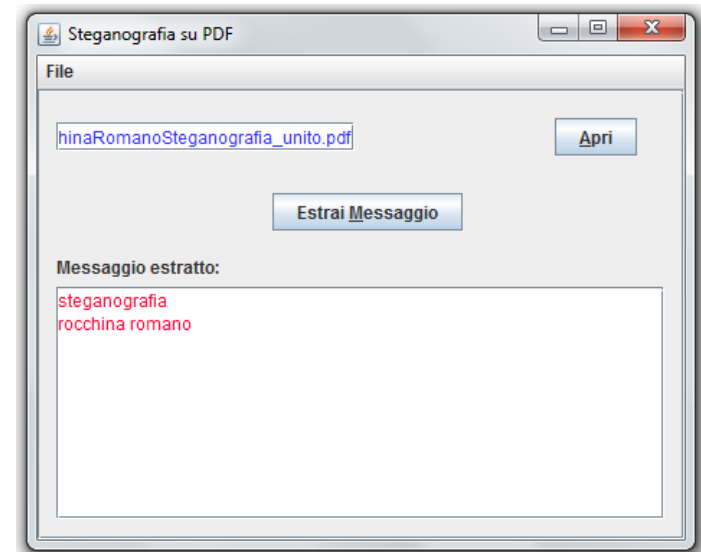


## RISULTATI RAGGIUNTI 2/3

- 3) il contenuto steganografico **non "sopravvive"**, se decidiamo di copiare **parte** dello stego PDF in un documento **MS-Word**;



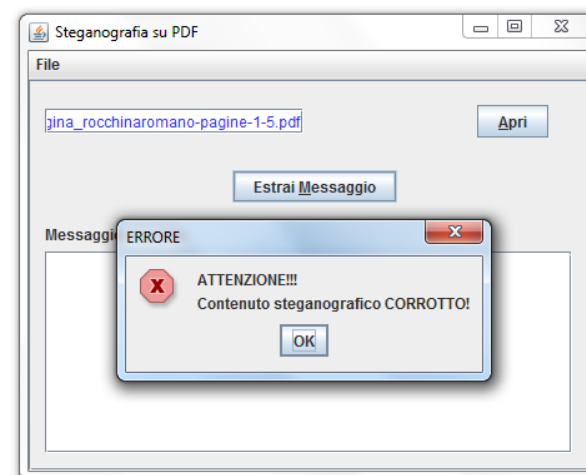
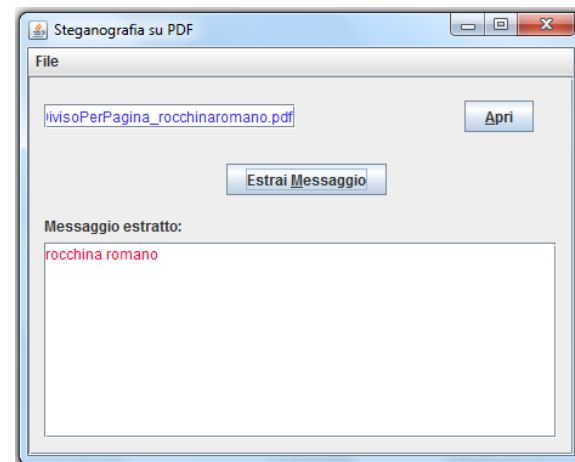
- 4) il contenuto steganografico **"sopravvive"**, se **uniamo** due o più **stego PDF**;



## RISULTATI RAGGIUNTI 3/3

5) Se **eliminiamo alcune pagine** dello stego PDF, invece, è possibile osservare che

- se il messaggio segreto inserito è **uguale per tutte le pagine** del documento PDF, il contenuto steganografico "**sopravvive**";
- al contrario, se durante il processo di embedding il messaggio segreto è stato **diviso per il numero di pagine** del documento PDF, il contenuto steganografico andrebbe **perso**.







# UNIVERSITA' DEGLI STUDI DELLA BASILICATA

Dipartimento di Matematica, Informatica ed Economia

Corso di Laurea in Scienze e Tecnologie Informatiche

---

## CONCLUSIONI E SVILUPPI FUTURI

1. **migliorare il tempo di esecuzione** per il processo di embedding;
  2. implementare il processo steganografico, ivi proposto, utilizzando un'**altra libreria** anche in un linguaggio di programmazione diverso da Java;
  3. per evitare che il **mouse possa rilevare la presenza** del contenuto steganografico, si potrebbe individuare l'area del documento PDF, in cui è inserito il testo, e fornirla come prima coordinata per l'inserimento delle immagini bianche;
  4. sviluppare un'applicazione per **dispositivi mobile** (iOS o Android);
  5. cambiare il **processo di codifica/decodifica** per l'inserimento e l'estrazione del messaggio segreto;
  6. trovare un possibile **attacco di steganalisi** per questo processo steganografico.
-



UNIVERSITA' DEGLI STUDI DELLA BASILICATA  
Dipartimento di Matematica, Informatica ed Economia  
Corso di Laurea in Scienze e Tecnologie Informatiche

---

## DISTRIBUZIONE DEL CODICE SORGENTE

➤ Il codice sorgente **Java**, utilizzato per implementare questo processo steganografico, è disponibile al seguente indirizzo:

<https://github.com/rocchinaRomano?tab=repositories>

➤ nella **repository** “[steganografiaSuPDF](#)”

➤ all'interno della quale è possibile trovare il file “[README.txt](#)”, che fornisce tutte le **informazioni necessarie** per il **download** e l'utilizzo dell'**applicazione** sviluppata.

➤ per ulteriori **info**, è possibile scrivere qui:

[rocchina.romano002@studenti.unibas.it](mailto:rocchina.romano002@studenti.unibas.it)

---

Grazie

a tutti

per l'attenzione!!! 😊