

Steganografia w HTML

Łukasz Polak



Plan prezentacji

- Co to jest steganografia? Historia i współczesność
- Rodzaje steganografii
- HTML – język znaczników
- Możliwości zastosowania steganografii w HTML

Steganografia

"Are you deaf, Father William!" the young man said,
"Did you hear what I told you just now?
"Excuse me for shouting! Don't waggle your head
"Like a blundering, sleepy old cow!
"A little maid dwelling in Wallington Town,
"Is my friend, so I beg to remark:
"Do you think she'd be pleased if a book were sent down
"Entitled 'The Hunt of the Snark?' –
"Pack it up in brown paper!" the old man cried,
"And seal it with olive-and-dove.
"I command you to do it!" he added with pride,
"Nor forget, my good fellow, to send her beside
"Easter Greetings, and give her my love."

Adelaide Paine

1876 – Lewis Carroll

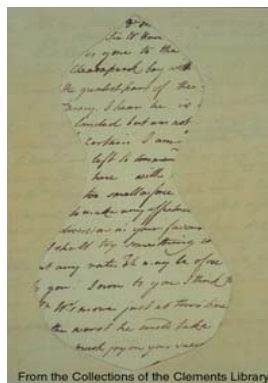
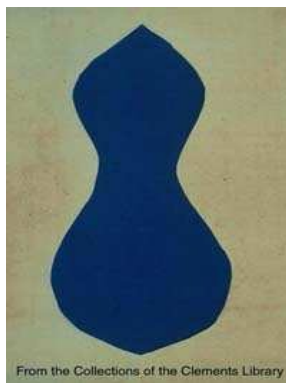
Steganografia:

- (gr. „ukryte pismo”) przesłanie informacji, w taki sposób, aby postronne osoby nie podejrzewały nawet jej istnienia

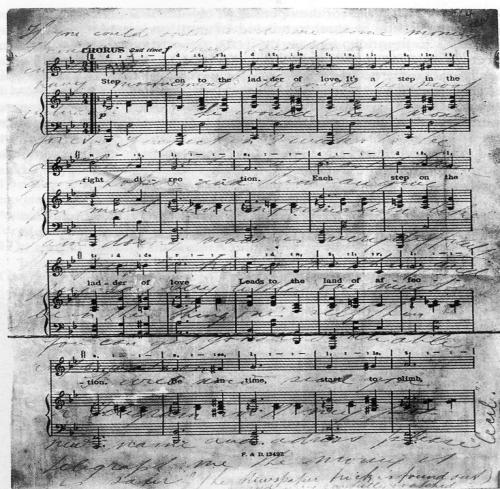
Pierwsze wzmianki – Herodot (V w.p.n.e.)

- Tatuowanie głowy niewolników
- Pisanie na drewnianych tabliczkach pokrywanych woskiem

Historia steganografii cd.



Maskowanie

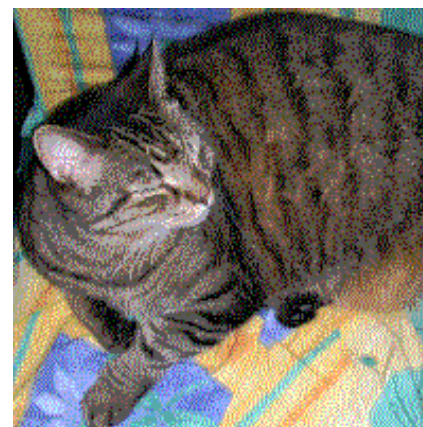


Atrament sympatyczny

Steganografia cyfrowa

Współczesna steganografia – możliwości zastosowania:

- Obrazy
- Dźwięk
- Video
- Tekst
- Pliki wykonywalne
- Steganografia sieciowa



Steganografia tekstowa

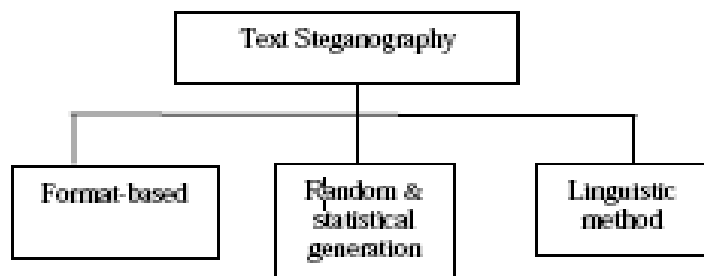


Fig.3: Three basic categories of text steganography

"To Anna jest naszą egzaminatorką"

Sposoby ukrycia informacji :

- Użycie wybranych znaków w tekście do przekazania informacji
- Formatowanie tekstu – białe znaki, odstępy między wierszami, wielkość czcionki, sposób pisania niektórych liter np. b,p,d,t
- Używanie metod semantycznych – synonimy, skróty
- Ukrywanie informacji w losowych ciągach znaków, słów – czy to wciąż steganografia?

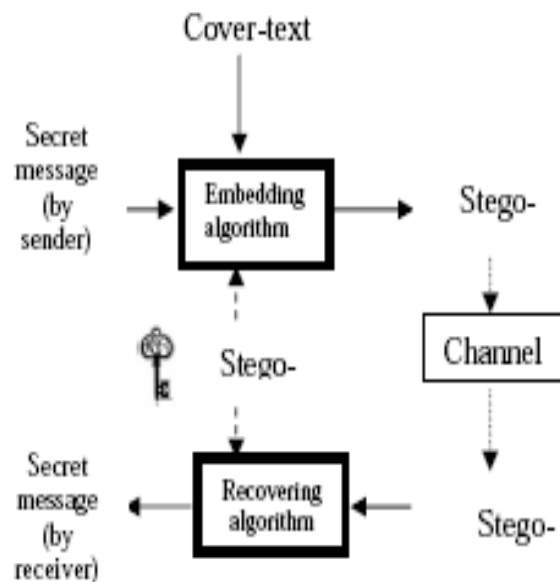


Fig. 2: The Mechanism of Text Steganography

SNOW

SNOW exploits the
steganographic
nature
of
whitespace



Projekt (od 2006 na licencji GPL), mający na celu ukrywanie wiadomości w plikach tekstowych poprzez dodawanie białych znaków na końcu linii.

www.darkside.com.au/snow/

HTML

Hypertext Markup Language

- najpopularniejszy język wykorzystywany do tworzenia stron WWW
- bazuje na znacznikach
- pierwsza specyfikacja: *HTML Tags* Berners-Lee (1991 r.). Zawierała 22 znaczniki, z czego 13 jest wykorzystywanych do dzisiaj
- obecna wersja – 4.01 (w styczniu bieżącego roku W3C opublikowało draft wersji 5)
- XHTML (2000r.)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
    "http://www.w3.org/TR/html4/strict.dtd">

<html lang="pl">

<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
  <title>Tytuł strony</title>
</head>

<body>
  <h1>Nagiówek pierwszego poziomu</h1>
  <p>
    Akapit (z ang. paragraph)<br>
    i jeszcze trochę tekstu w następnym wierszu<br>
    <a href="http://pl.wikipedia.org">
      a to link do Wikipedii</a>.
  </body>

</html>
```


Steganografia w HTML (1)

Tradycyjne metody tekstowe:

- Używanie białych znaków w tekście

Efektywniejsza metoda:

- Użycie białych znaków na końcu znacznika

<tag> -> 0

<tag/ > -> 1

```
<p>Steganografia</p>
```

```
<p >Steganografia</p>
```

10

```
<p >Steganografia</p >
```

11

Alternatywna metoda - zmienianie wielkość liter znaczników, żeby zakodować informację (np. <tAG> -> 011).

Wady tej metody:

- łatwa do wykrycia
- tylko czysty HTML

Steganografia w HTML (2)

Używanie domyślnych wartości atrybutów

```
<HR width=„100%”>
```

0

```
<HR>
```

1

Trudniejsza do wykrycia niż poprzednio omówione metody

Steganografia w HTML (3)

Zmianianie kolejności atrybutów

```
<IMG SRC="picture.jpg" ALT="A picture">
```

0

```
<ALT="A picture" IMG SRC="picture.jpg" >
```

1

Przy 8 atrybutach mamy 40320 (8!) kombinacji co pozwala na zakodowanie prawie 2 bajtów informacji (15,3 bit)

Zalety:

- nie wpływa na rozmiar pliku
- trudno wykrywalna

Implementacja - DEOGOL

Deogol w staro-angielskim znaczy „ukryty”.

Site	Capacity (bytes)
news.google.com	346
www.netscape.com	324
www.cbc.ca	212
www.cnn.com	183
www.microsoft.com	119
www.uwaterloo.ca	116
slashdot.org	127
www.theonion.com	104
www.ebay.com	76
www.memepool.com	59
www.google.com	18

Efektywność DEOGOLa. Dane na 10/2002.

<http://wandership.ca/projects/deogol/intro.html>

Steganografia w XHTML/XML

XHTML a w szczególności XML, oferują dodatkowe możliwości ukrywania informacji (m. in.):

- wykorzystanie dwojakej reprezentacji pustych znaczników (lub)
- zmienianie kolejności występowania elementów (tylko XML)

Podsumowanie

- HTML jako strukturalny język znaczników daje duże możliwości przekazywania ukrytych informacji
- HTML jako medium do przenoszenia ukrytych informacji jest znacznie mniej pojemne od obrazów i plików audio

- Steganografii tekstowej poświęcono dotychczas znacznie mniejszą uwagę niż innym jej rodzajom.
- Istnieją tylko nieliczne prace na temat ukrywania informacji w ustrukturyzowanych dokumentach – (X)HTML, XML
- Praktycznie nie istnieją narzędzia do analizy kodu HTML, pod kątem ukrytych informacji
- HTML jest dzisiaj wszechobecny
- XML jest coraz popularniejszym, uniwersalnym językiem do wymiany danych

Podsumowanie cd.

Co daje nam steganografia HTML?

- Łatwa wymiana ukrytych informacji
- Bezpieczna – bardzo małe ryzyko ujawnienia faktu istnienia ukrytej informacji

Wniosek: Steganografia HTML/XML jest warta naszej uwagi 😊



Koniec

Dziękuję za uwagę