Natural Language Watermarking

Mercan Topkara

Center for Education and Research in Information Assurance

Department of Computer Sciences

Cuneyt Taskiran

Edward J. Delp

Video and Image Processing Laboratory

School of Electrical and Computer Engineering

CERIAS

PURDUE UNIVERSITY

VIPER

Problem and Key Idea

Performing natural language watermarking, which uses the structure of the sentence constituents in natural language text to insert a watermark.

Outline

- **Why Natural Language (NL) watermarking?**
- **NL Watermarking vs. Image Watermarking**
- **Natural Language Processing (NLP) for Watermarking**
- **Linguistic Steganography**
- **NL Watermarking**
- **Conclusions**

Why Natural Language Watermarking?

- **Authenticating the source of a document**
- **Proving or denying ownership on a document**
- **Controlling distribution and reuse of intellectual property**
- **Digital libraries, on-line news channels, online stores etc.**
- **Content protection, text auditing, meta-data binding, tamper-proofing, traitor tracing**

Natural Language vs. Image Watermarking

- **Same goal, different methods**
- **The amount of redundancy is very low in text**
- **Evaluation of stealthiness is harder for text**
- **Limitations of human visual system is high**
  - LSB vs. Synonym Substitution

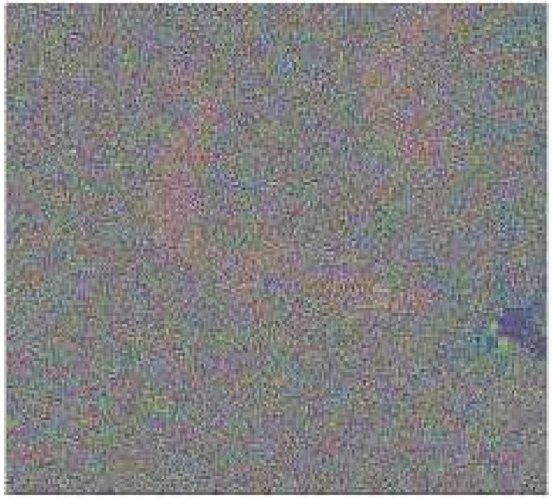
Natural Language vs. Image Watermarking

LSB Embedding

original

watermarked

difference image



Synonym Substitution

“The tyrant will always find a pretext for his tyranny.”

“The despot will always find a pretext for his despotism”.

Natural Language vs. Image Watermarking

- **Sentences have combinatorial syntax and semantics:**
  - complex representations are constructed using structurally simple constituents
  - the semantic content of a sentence is a function of the semantic content of its constituents together with its syntactic structure
- **Transformational Grammar Theory**

“Ned loves Jody” ↔ “Jody is loved by Ned”

- **Not the surface but the underlying structure is altered**
  - a rough analogy to modifying DCT coefficients
- **Achieving higher robustness**

NLP for Watermarking

- Natural Language Processing (NLP) aims to design algorithms that will analyze, understand, and generate natural language automatically
- Electronic Data Resources and Tools
  - Corpora
  - Dictionaries e.g., WordNet, Verbnet
  - Parsers, Generators, Machine Translation and Question Answering Systems

NLP for Watermarking: Linguistic Transformations

- Synonym Substitution
- Syntactic Transformations
- Semantic Transformations

NLP for Watermarking:
Syntactic Transformations

Transformation	Original sentence	Transformed sentence
Passivization	The slobbering dog kissed the big boy.	The big boy was kissed by the slobbering dog.
Topicalization	I like bagels.	Bagels, I like.
Clefting	He bought a brand new car.	It was a brand new car that he bought.
Preposing	I like big bowls of beans.	Big bowls of beans are what I like.
There-construction	A unicorn is in the garden.	There is a unicorn in the garden.
Fronting	“What!” Alice cried.	“What!” cried Alice.

NLP for Watermarking:
Syntactic Transformations

- **Verb Alternations:**
  - **Levin Verb Classes**
  - **[Spray/Load Alternation]**

"Jack sprayed paint on the wall. ⇔
Jack sprayed the wall with paint."

"The farmer loaded apples into the cart . ⇔
The farmer loaded the cart with apples ."

NLP for Watermarking: Semantic Transformations

- Based on co-references
  - **Pruning :** removing repeated information

    Yet Iceland has offered a residency visa to ~~ex-chess champion~~ Bobby Fischer in recognition of a 30-year-old match that put the country on the map.
  - **Grafting:** adding or repeating information

    He, an American citizen, is being detained in Japan and is wanted in the US for violating international sanctions against the former Yugoslavia by playing there in 1992.
  - **Substitution:** replacing information

    Ex-chess champion's historic win over Russian Boris Spassky in Reykjavik in 1972 shone the international spotlight on Iceland as never before.

NLP for Watermarking

- **NL Parsing**
  - **processing sentences to determine their structure**

Ned loves Jody. ↔ Penn TreeBank Structure

```
(S (NP Ned)
      (VP loves (NP Jody))
 (. .))
```

- **NL Generation**
  - **constructing NL output from non-linguistic information representations according to some communication specifications**

Sentence realization

```
DSYNTS:
love [ class:"verb" ]
( I Ned   [ class:"proper_noun" ]
  II Jody [ class:"proper_noun" ]
)
END:
```

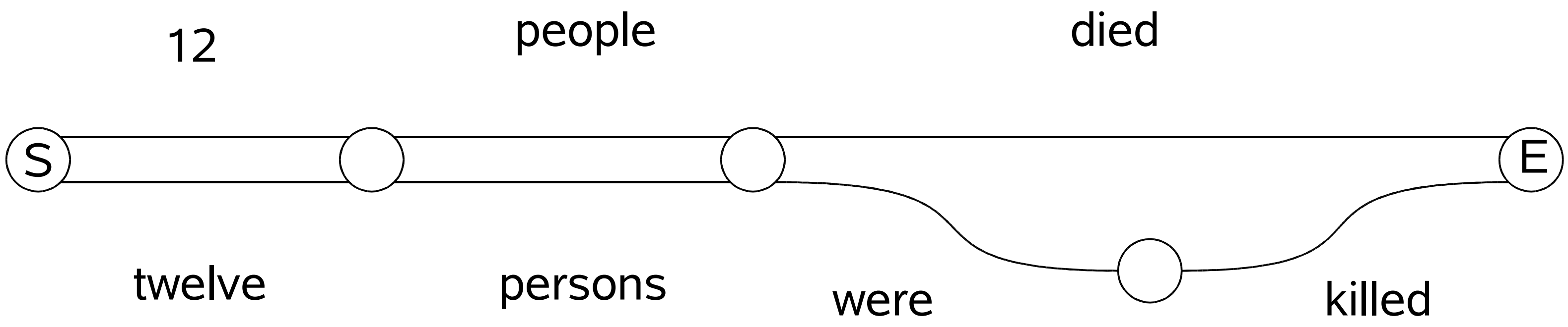
NLP for Watermarking

- **Paraphrasing**
  - **Parallel corpus**

    After the latest Fed rate cut, stocks rose across the board.

    Winners strongly outpaced losers after Greenspan cut interest rates again.

  - **Finite-State Automata Reduction**



Previous Work in Linguistic Steganography

- **Mimicry Text: Using PCFGs to Generate Cover Text**

(Wayner, 1992)

Rule	Code	Prob
S =>AB	0	0.5
S =>CB	**1**	**0.5**
A =>She	00	0.25
A =>He	01	0.25
A =>Susan	10	0.25
A =>Alex	11	0.25
B =>likes D	0	0.5
B =>detests D	10	0.25
B =>wants D	**110**	**0.125**
B =>hates D	111	0.125
C =>Everybody B	**0**	**0.5**
C =>The lady B	10	0.25
C =>A nice kid B	11	0.25
D =>milk.	00	0.25
D =>apples.	01	0.25
D =>pie.	**10**	**0.25**

Position	Prefix	Output
°1011010	1	CB
1°011010	0	Everybody B
10°11010	110	Everybody wants D
10110°10	10	Everybody wants pie.

Previous Work in Linguistic Steganography

- **Synonym Substitution: Using mixed radix form**

(Winstein, 1999)

	0 wonderful		
	1 decent		0 city.
Midshire is a	2 fine	little	
	3 great		1 town.
	4 nice		

$$(101)_2 = 5$$

$$\begin{pmatrix} a_1 & a_0 \\ 5 & 2 \end{pmatrix} = 2a_1 + 1a_0 = 5$$

- **NICETEXT: Using a dictionary table and style template**

(Chapman and Davida, 97)

Type	Code	Word
name–male	0	Ned
name–male	1	Tom
name–female	0	Jody
name–female	0	Tracy

Style

name–female name–male name–male

Payload

011

Cover Text

jody tom tom

Previous Work in NL Watermarking

- **Synonym Substitution (Atallah et al., 2000)**
  - If $m_i \bmod k = 1$ and $A(w_i) + r_i \bmod k$ is a quadratic residue modulo *p*, then w_i is kept same.
  - $m_i \bmod k$ is the current bit of watermark
  - $A(w_i)$ is the ASCII value of word w_i
  - *p* is a 20 digit prime key
  - *k* is the number of bits in the watermark
  - $r_0, r_1, \ldots r_{k-1}$ is the sequence of pseudo-random numbers generated using *p* as key

Previous Work in NL Watermarking

- **Using sentence tree structure (Atallah et al., 2001, 2002)**
  - DCT analogy
  - Selection depends on the tree structure
  - Nodes are labeled in pre-order traversal
  - Then a node label j is converted to 1 if $j+H(p)$ is a quadratic residue modulo *p*
  - B_i is generated according to post-order traversal
  - A rank d_i is assigned for each sentence s_i using
    $$d_i = H(B_i) \oplus H(p)$$
  - Starting from least-ranked sentence s_j the watermark is inserted s_j's successor in the text by altering B_{j+1} using linguistic transformations

Previous Work in NL Watermarking

- **With Syntactic Transformations**

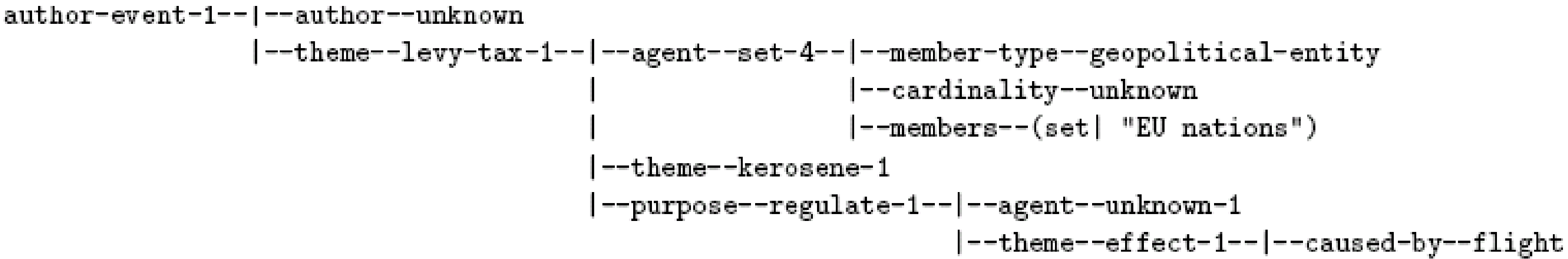
(Atallah et. al, 2001)

(S (NP Ned)
(VP loves (NP Jody))
(. .))

- **With Semantic Transformations**

(Atallah et al., 2002)

The EU ministers will tax aviation fuel as a way of curbing the environmental impact of air travel.



```
author-event-1--|--author--unknown
                |--theme--levy-tax-1--|--agent--set-4--|--member-type--geopolitical-entity
                                      |                |--cardinality--unknown
                                      |                |--members--(set| "EU nations")
                                      |--theme--kerosene-1
                                      |--purpose--regulate-1--|--agent--unknown-1
                                                              |--theme--effect-1--|--caused-by--flight
```

Summary

- Steganography vs. Watermarking
  - More complex methods
- Still in its infancy
- Gain pace with synergy
- Statistical methods for robust and generic solutions
- Easier to work on syntactic structure
  - Available tools are better developed
- Harder to Evaluate
  - Different styles, genres, authors and audience

Conclusion

- **Challenging problem !**
- **Collaboration with image watermarking will help**
- **Existing work is on preliminary level**
- **Wide range of application areas**