

Lecture 4. Stegosystems for other cover objects.

4.1. *Linguistic* SG (SG-L)

4.2. *Graphic* SG (SG-G)

4.3. *Internet (Network)* SG (SG-I)

Some of SG proposed on Internet

Type of SG	Operation System	Method of embedding	Cover objects	Amount of embedded information	Additional security measures	Remarks
Steganos Security Suite 2006	Windows	?	Moveless images and musical files	?	Password, Encryption of messages	Using USB as the key. It is possible to erase data if the key is lost or stolen.
StegoVideo	Windows	?	Video files	?	Password	It is possible to compress video files after embedding.
StegaNote	Windows	Modified LSB embedding	Moveless images of BMP format 24 bit/pixel	Maximum data embedding rate 1/255	?	Friendly interface
StegoMagic	Windows	?	Text files, Formats WAV, BMP 24 bits/pixel	Maximum data embedding rate 1/8	Password, Encryption of messages with DES.	Friendly interface
Puff	Windows	16 different embedding algorithms	Formats BMP, JPG, PCX, PNG, TGA, AIFF, MP3, NEXT / BC, WAV.	?	?	?
wbStego4.3open	Windows	?	Formats BMP, TXT, HTML / XML, PDF	Up to 2 Gb information	Encryption of messages	Friendly interface
Steganography 4.0	Windows	?	Format BMP	?	Encryption of messages	Friendly interface

SecurEngine Professional 1.0	Windows	?	Formats: BMP, GIF, PNG, HTM.	?	Encryption of messages with AES, Blowfish, GOST, Triple-DES	?
Hermetic Stego v6.5	Windows	?	Format BMP	?	?	?
PhotoCrypt 1.1	Windows	?	Format BMP	Up to 50% of CM	?	?
Invisible Secrets v4.0	Windows	?	Formats: JPEG, PNG, BMP, HTML and WAV	?	Encryption with Blowfish, Twofish, RC4, Cast128, and GOST	There is a password generator.
CryptArkan	Windows	?	Formats: WAV and BMP.	?	Information is encrypted	?
Gifshuffle v2.0	Windows	Transposition of colours according of given colourmap).	Format GIF	?	Information is encrypted	Image compression is possible
JPegX	Windows	?	Format JPEG.	?	Information is encrypted with password.	?
The Third Eye	Windows	?	Formats: BMP, GIF and PCX.	?	Information is encrypted	Friendly interface

WeavWav	Windows	?	Format WAV	?	?	Friendly interface.
InfoStego	Windows	?	Format BMP	?	Information is encrypted	Image compression is possible
Camouflage	Windows	Attaching of encrypted (scrambled) file to cover	Any file format	?	Information is encrypted with password.	Embedded information is detected very simple but it cannot be read
BMP Secrets	Windows	?	Format BMP	Up to 65% cover message size	Information is encrypted	?
S-Mail Shareware v1.3	Windows и DOS	?	Formats EXE and DLL	?	Information is encrypted	It is secure against simple methods
S-Tools v4	Win 95/ NT	?	Formats :BMP, GIF and WAV	?	?	?
Encrypt Pic	Windows	?	Format BMP 24 bits/pixel	?	Information is encrypted with Blowfish.	?
Contraband Hell Edition	Windows	?	Format BMP		Information is encrypted	Friendly interface.

Steghide 0.4.6.b	Windows и Linux	?	Formats : JPG, BMP, WAV .	?	Information is encrypted with Blowfish, pseudorandom embedding into cover; 128 MD5 for the key to Blowfish.	?
Hide4PGP v2.0	Windows, DOS, OS / 2, and Linux	?	Formats : BMP, WAV and VOC	?	?	?
Blindside	Windows, Linux, HP, Solaris and AIX	?	Format BMP	?	Information is encrypted	?
TextHide	Windows	Change words into the text by their synonymous	Text files	?	Information is encrypted with Twofish .	Text files are in
JP Hide and Seek	Win95/98/ NT, DOS and Linux	?	Format JPG	?	Information is encrypted with Blowfish.	Small software size
MP3Stego	Windows 95/98/NT иandLinu x / Unix	?	Formats : MP3 and WAV.	?	?	Files can be compressed with MPEG ⁵

Stella	?	Two embedding algorithms.	Formats : GIF, BMP and JPEG.	?	Information is encrypted	?
SGPO	?	Transposition of colours according of given colourmap).	Format GIF	?	?	?
Snow	?	The use of tabulation and spaces at the end of rows	Text files	?	Information is encrypted with ICE.	Compression is possible
Invisible Encryption (IVE)	?	?	Format GIF	?	?	The use of passwords.
Visual Encryption (VE)	?	?	Format GIF	?	?	The use of passwords.

We can see that there are a lot of proposed SG for different CO but undetectability of them is not granted.

4.1. Linguistic SGS.

Definition: Embedding of any data into text documents that is prepared in any natural language.

The main requirement: SG-L has to be appear as innocent text .This means that content, grammar, syntax and semantic should be kept completely.

Two main types of SG-L:

1. With given CO (text) .
2. With chosen CO (text).

Basic principle to design SG-L of the first type.

To find the areas with uniformly distributed data and replace them to the encrypted messages .

SG based on substitution of absolute and relative synonyms.

Definitions.

1. *Synonyms* are words that can replace each other in some class of contexts with insignificant change of the whole text meaning. The reference “some class” and to “insignificant change “make this definition rather vague , nevertheless nearly all modern synonymy dictionaries are based on it.

2.Absolute synonyms.

These are absolute synonyms that can replace each other in any context without any change in meaning.

Examples : sofa – settee, big – large, another – different, mind – opinion
United States of America-USA-US-Unites States , former President-ex-President, stegosystem- SG-system .

3. *Relative synonyms*. These are synonyms that can replace (or not) each other depending on context .

Examples :

client -consumer -customer -user,
come -go - leave , chief - leader -head
real number – continuous number, real life ≠ continuous life.

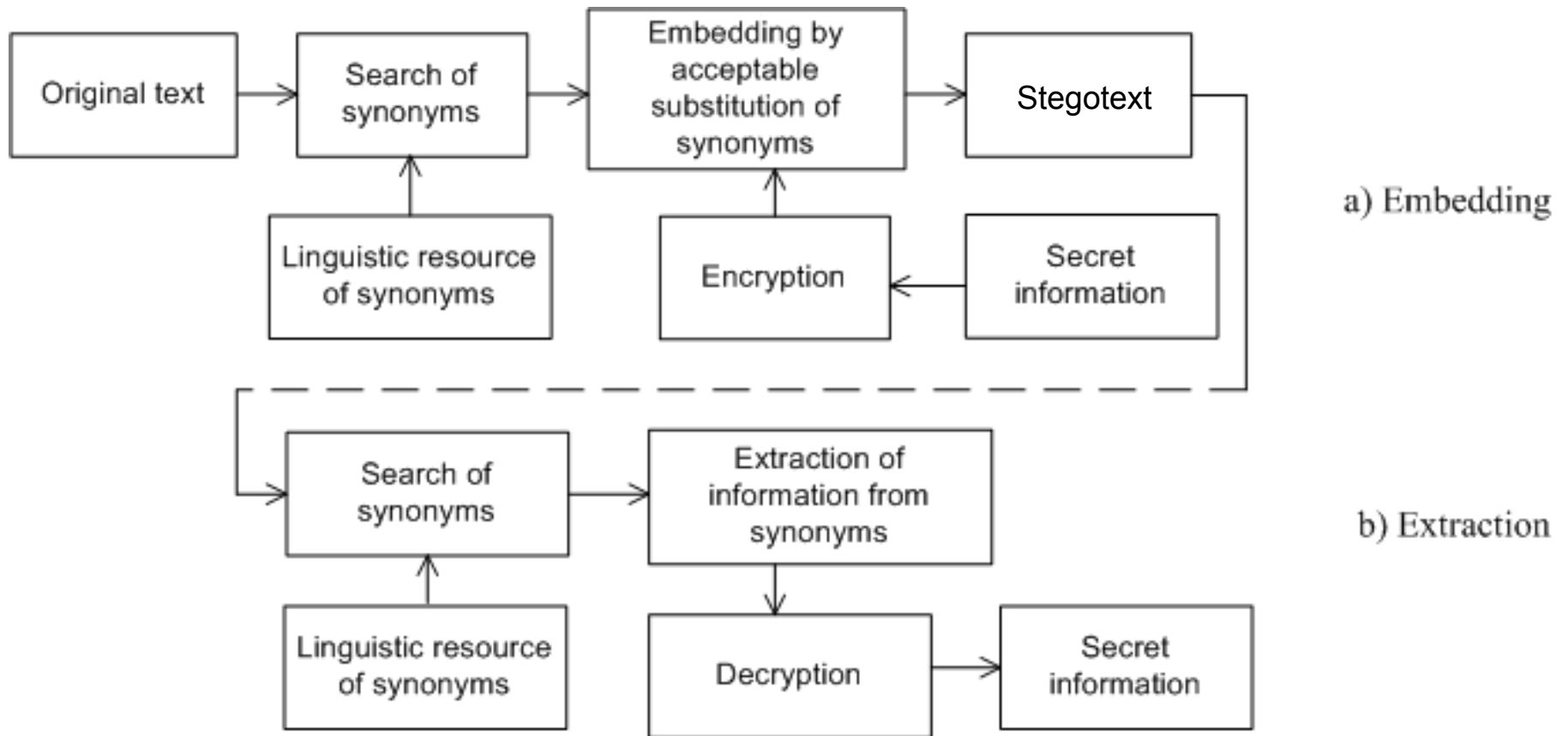
Both absolute and relative synonyms are collected in special dictionaries for the thing:

1. “Dictionary of Russian Language Synonyms” , St.Petersburg, 2006 (in Russian).
2. Oxford Collocation Dictionary for Students of English. Oxford University Press. 2003.
3. Fellbaum Ch. WordNet: An electronic lexical database. MIT Press, 1998.

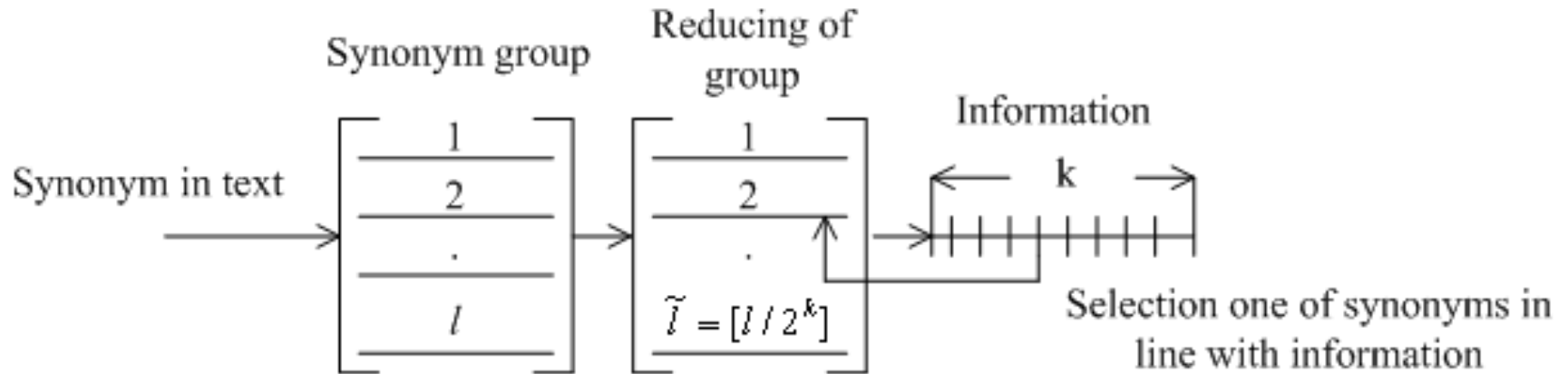
Remark .Due to N. Chomsky collocations are simply a series of two or more words occurring together in a narrow window moving along a text.

Examples : play -> the role , new <- method.

Embedding and extraction algorithms for SG-L based on synonyms.



Embedding into synonymy groups.



Examples:

----- → 00 ----- → 0
----- → 01 ----- → 1
----- → 10
----- → 11

(There may be more effective methods of encoding).

Example of synonyms-based SG-L design.

Cover text :

As many as five **subterranean pulses** are *registered* during **24 hours** in the south of *Altai Republic*. The *strength* of the **earthquakes** *amounts* from 2.2 to 3.1 points on Richter scale , as they have *informed* in the Aktash seismic station today *in the afternoon*.

(Absolute synonyms highlighted with bold face , whereas relative synonyms are marked by cursive lines).

Encoding of absolute synonyms:

earthquakes (0) – subterranean pulses (1),
one day (0) – during 24 hours (1),

Encoding of relative synonyms:

registered (00),
fixed (01),
marked(10),
remarked (11),
Altai Republic (0) – Altai (1),
amounts (0) – was (1),
informed (0) – communicated (1),
afternoon (0) – second part of the day (1),
strength (00), amplitude (01), magnitude (10), power (11).

After checking of all synonymy groups on compatibility with their “surrounding” it is performed a substitution of these synonyms given information to be embedded.

Total number of bits that can be embedded in this fragment is 10.

Investigation of LS-L shows that steganographic bandwidth (SB) or in another words –data embedding rate [2] is about 0,004 .This means that cover text should be 250 times longer than the hidden information.

Another method of design LS-L: A changing of the word order in the sentences.

Example: New earthquake has taken place in Iran on Monday_.

 S V L T

L –where , T – when, V – verb, S – subject.

There exists $4! = 24$ permutation totally but not all of them are acceptable in line with grammar of some language. So for Russian language translation of this sentence are acceptable only 4 transpositions:

TLVS, SVTL, TVSL , LTVS .

But in English the word order is very strong and the number of transpositions will be smaller.

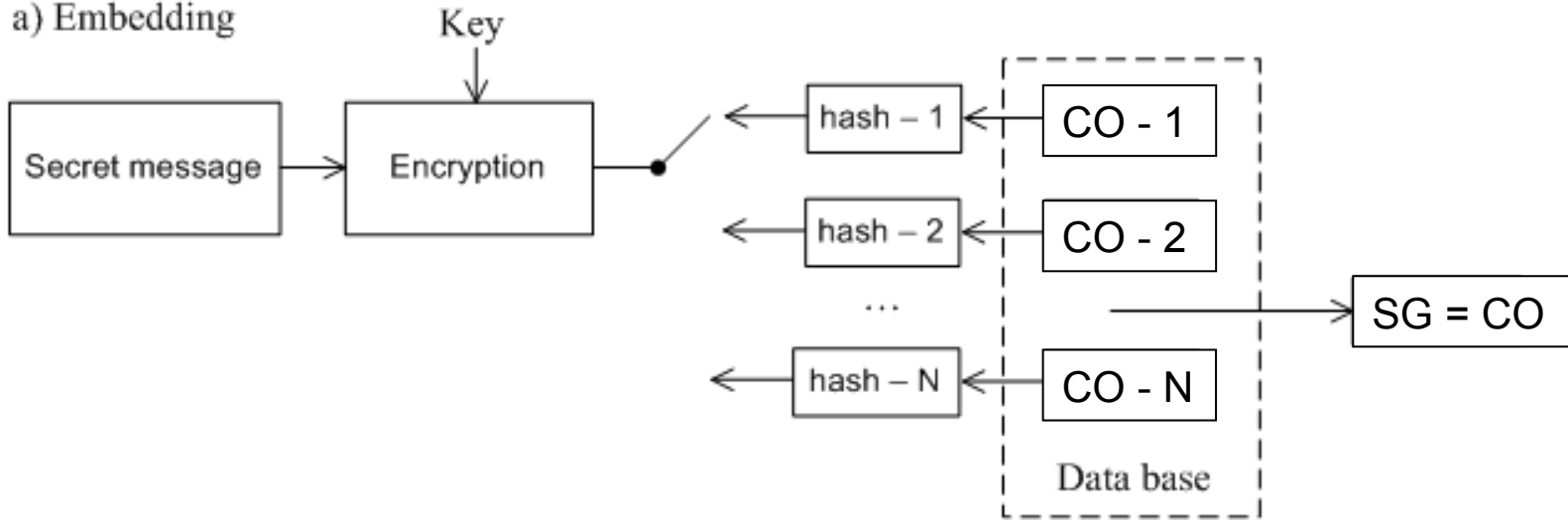
Let us estimate potential SB using Meaning-Text Theory. The French sentence in 35 words reveals 50 million synonymous variants [3]. It means that each paraphrase of the sentence can hide in itself 25.6 bits of information, this giving SB of about 0.016.

The main problem for implementation of this method is the limited availability of the abovementioned linguistic resources.

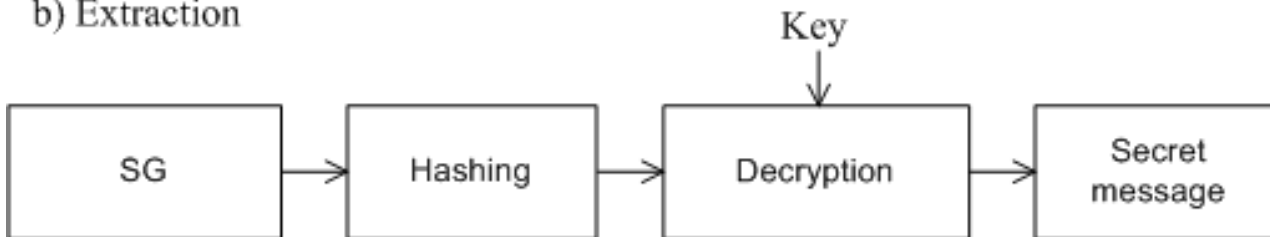
2. SG-L with chosen (edited) CT (cover text) .

It is in fact a particular case of the following general ideal SG for any cover messages:

a) Embedding

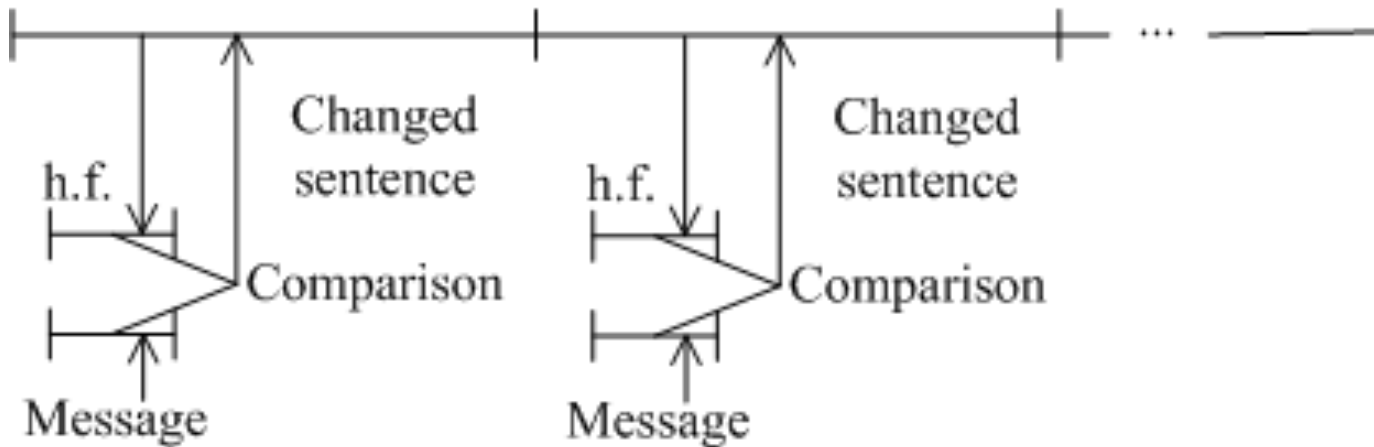


b) Extraction



Remark. In this SG hash function and data base of CO is public but only encryption/decryption key is kept in secret that plays a role of stegokey.

Implementation for the case of SG-L:



Properties of all SG-L:

1. Ideal security.
2. Any CT can be used for embedding.
3. Low BD (embedding rate).
4. A vulnerability to “blind” removal attack of the embedded information that can be performed by reembedding of truly random messages into stegotext.
5. SG-L with edited cover text does not depend on language and does not require linguistic resources.

4.2. Graphic (raster) SG (SG-G).

CO – graphic (image) document (text, picture, scheme, table, ctr.)

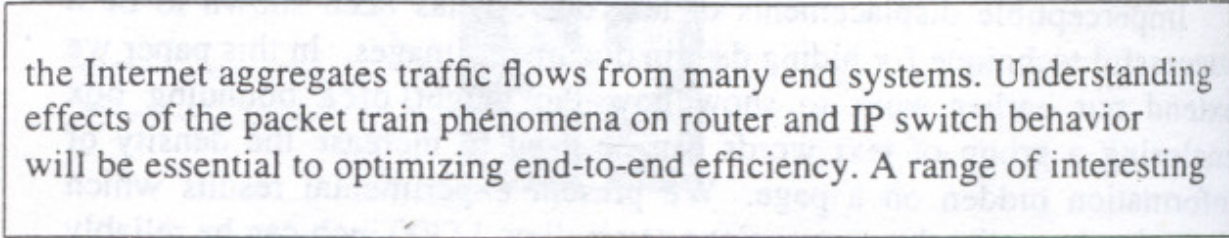
The simplest methods of embedding into image-text-documents:

- changing spaces between words and or sentences,
- changing spaces between rows ,
- lifting words or rows up and down,
- small rotation of rows or wordse.

(See a demonstration on the next slide.)

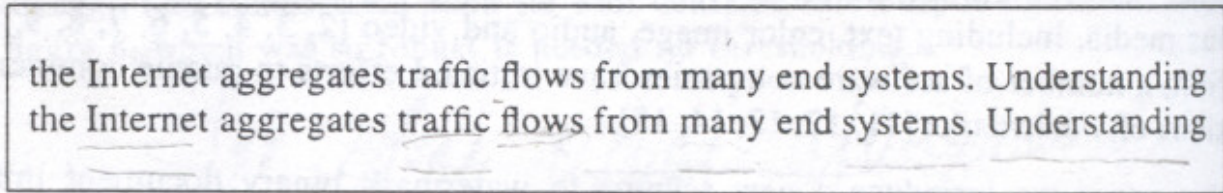
All these SG-G can be easily detected by the use of statistical steganalysis.

Examples of embedding in text files [4]:



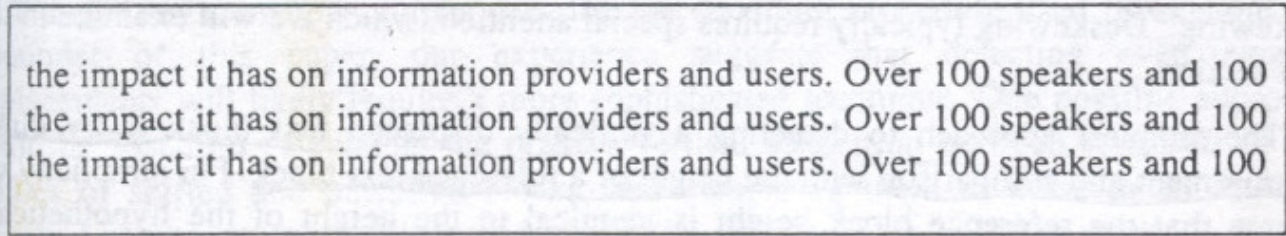
the Internet aggregates traffic flows from many end systems. Understanding effects of the packet train phenomena on router and IP switch behavior will be essential to optimizing end-to-end efficiency. A range of interesting

Figure 1 - Vertical shifting of a text line. The first and third lines are unshifted; the second line has been shifted by $1/300$ inch. Can you tell if it has been moved up or down?



the Internet aggregates traffic flows from many end systems. Understanding the Internet aggregates traffic flows from many end systems. Understanding

Figure 2 - Horizontal shifting of words on a text line. The first contains no shifted words; on the second line the 2nd, 4th, 6th and 8th words are each horizontally displaced by $1/300$ inch. Line length remains unchanged.



the impact it has on information providers and users. Over 100 speakers and 100
the impact it has on information providers and users. Over 100 speakers and 100
the impact it has on information providers and users. Over 100 speakers and 100

Figure 3 - Illustration of marks inserted by lifting words off the baseline. The first line contains no shifted words; the second and third lines contain 3 words each shifted by $1/600$ and $1/300$ inch, respectively.

More sophisticated method “Simulation of scanner’s noises” (SG-S).

The main idea: Scan printed document and embed in it secret information simulating scanner’s noises.

Algorithm of embedding:

1. After scanning black and white document is divided consecutively into areas $n \times n$ pixels denoted by A .

Introduce the following notations: m –the number of black pixels in A , $m = m_+$ if A contains even number of black pixels, $m = m_-$ if A contains odd number of black pixels, $0 < k < \frac{1}{2}$ is chosen threshold, $b = \{0, 1\}$ is bit of secret information that has to be embedded in the area A , $A = A_0$, if $kn^2 < m < (1-k)n^2$, $A = A_1$, if $m = (1-k)n^2$, $A = A_2$, if $m = kn^2$.

2. If $A = A_0$, then embedding is determined by the following Table:

	$m = m_+$	$m = m_-$
$b = 0$	Change nothing	Change color of any one pixel to opposite value
$b = 1$	Change color of any one pixel to opposite value	Change nothing

Remark . It is possible to change any pixels but resting on a boundary between black and white .

3. If $A = A_1$, then embedding is determined by Table:

	$m = m_+$	$m = m_-$
$b = 0$	Change nothing	Change one of black pixels to white
$b = 1$	Change one of black pixels to white	Change nothing

4. If $A = A_2$, then embedding is determined by Table:

	$m = m_+$	$m = m_-$
$b = 0$	Change nothing	Change one of white pixels to black
$b = 1$	Change one of white pixels to black	Change nothing

5. If $A \neq A_0$, $A \neq A_1$, $A \neq A_2$ then nothing is embedded into this area .

Algorithm of extraction :

1. Divide the image consecutively into A -areas $n \times n$ pixels each.
2. If $A=A_0$, or $A=A_1$, or $A=A_2$, then extract $b=0$, if $m=m_+$ and $b=1$, if $m=m_-$.
3. If $A \neq A_0$, $A \neq A_1$, $A \neq A_2$, then extract nothing bits .

The main properties of SG-S above:

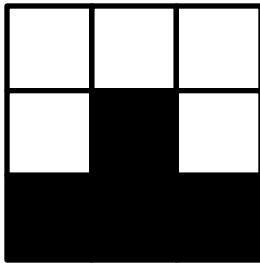
1. Extraction results in error free information.
2. The more are n and k , the more secure is SG but the less is embedding rate and vice versa .
3. SG is resistant to visual attack.
4. The embedded information can be removed easily by randomization of m_+ , m_- without significant degradation of cover image.

Attacks on Stegosystems

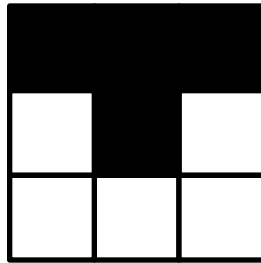
- Analysis of the number of single rejections
- Analysis of the number of single deepening's

Attack based on the number of single rejections

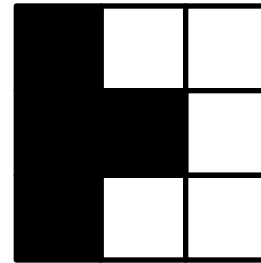
Single rejections



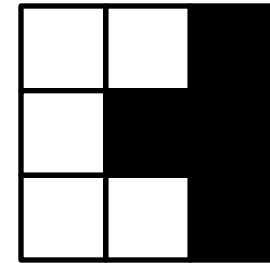
Up



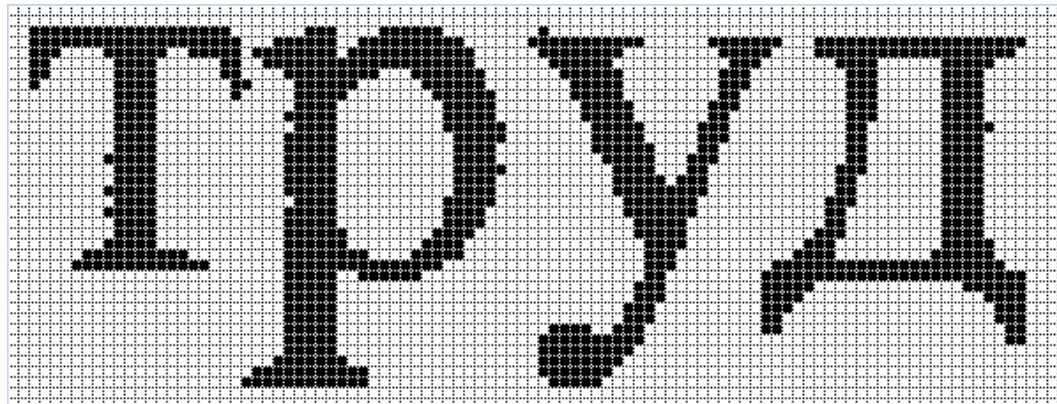
Down



Right



Left



Attack based on the number of single rejections

Hypothesis : the same amount of text on a page of A4 on average have less number of single rejections, than after the embedding

Estimated threshold : number of single rejections for different amounts of text

Analysis of the number of single rejections

Img. №	Up	Down	Left	Right	All before	Up	Down	Left	Right	All after
1	462	509	495	492	1958	602	542	649	720	2513
2	545	609	607	672	2433	654	654	757	936	3001
3	601	660	695	555	2511	743	713	866	787	3109
4	637	701	694	712	2744	788	773	851	951	3363
5	617	717	623	673	2630	799	791	806	887	3283
6	661	704	625	587	2577	818	770	787	825	3200
7	607	678	725	651	2661	750	735	862	903	3250
8	594	791	675	660	2720	743	866	819	897	3325
9	586	671	725	663	2645	728	728	897	881	3234
10	554	632	616	592	2394	691	708	761	815	2975
11	612	772	781	680	2845	782	830	937	915	3464
12	560	676	625	608	2469	696	726	788	823	3033
13	627	721	672	670	2690	746	776	782	885	3189
14	616	721	667	666	2670	780	767	845	872	3264
15	444	559	512	465	1980	565	622	621	666	2474
16	560	649	607	539	2355	693	695	778	764	2930
17	603	595	644	601	2443	734	655	788	808	2985
18	511	652	531	531	2225	591	705	682	742	2720
19	533	721	587	564	2405	665	782	748	790	2985
20	537	661	565	540	2303	679	716	702	792	2889

Attack based on the number of single rejections

The limits of applicability

- All text documents are printed on the same printer;
- All printed documents are scanned on the same scanner;
- Need database of test images to gather statistics;

Detection algorithm

- Thresholds are selected on the basis of the collected statistics depending on the density of the text on the page;

As a criterion for determining the density of the text on the page is used the number of black pixels on the page.

- Searching and counting single emission in the scanned text document;
- Counting the number of black pixels in the scanned document;
- Comparing counted single emissions with chosen thresholds;
- A decision is made whether image is cover-image or stego-image.

Testing the effectiveness of analysis based on single rejections

1. The following thresholds are selected based on analysis of 20 test images:
2. Hidden information is embedded with different speeds of embedding in 15 of 60 images submitted for stegoanalysis

Number of black pixels	Chosen threshold
600000 – 650000	1950
650000 – 700000	2150
700000 – 750000	2350
750000 – 800000	2550
800000 – 850000	2750
850000 – 900000	2950
900000 – 950000	3150

Testing the effectiveness of analysis based on single rejections

Cover – image №	Number of embedded bits
<i>Embedding 8000 – 12000 bits (n = 20 , k = 0.01)</i>	
45	11220
51	9400
57	9644
73	9978
79	10275
<i>Embedding 600 – 700 bits (n = 100 , k = 0.01)</i>	
23	608
27	581
42	621
48	640
50	618
<i>Embedding 400 – 500 bits (n = 100 , k = 0.1)</i>	
41	519
56	453
62	391
67	497
80	519

Testing the effectiveness of analysis based on single rejections

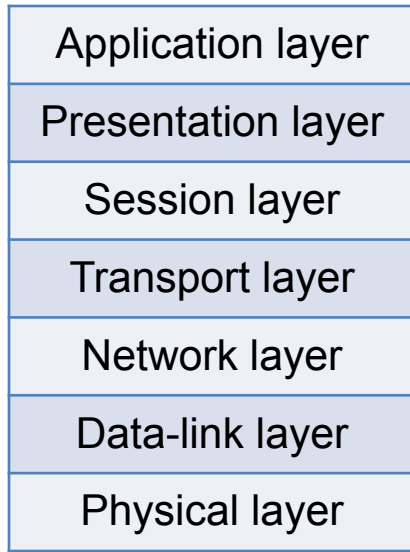
Image №	Up	Down	Left	Right	All	Number of black pixels	Detecting
23	466	653	632	712	2463	798781	Missed
27	439	563	653	631	2286	804339	Missed
41	501	669	753	718	2641	877083	Missed
42	508	659	733	698	2598	851749	Missed
45	795	775	998	1091	3659	946030	Stego - Image
48	508	684	813	710	2715	885834	Missed
50	507	644	759	747	2657	850749	Missed
51	601	719	850	937	3107	810658	Stego - Image
56	505	660	713	709	2587	813280	Missed
57	672	708	890	940	3210	829337	Stego - Image
62	467	560	676	661	2364	754718	Missed
67	499	666	825	791	2781	874827	Missed
73	740	748	992	1079	3559	859889	Stego - Image
75	531	655	800	795	2781	837569	False alarm
76	485	564	758	707	2514	742726	False alarm
79	718	748	1007	1128	3601	895940	Stego - Image
80	591	754	895	823	3063	907919	Missed

Testing the effectiveness of attack based on single emissions

Embedding speed, bits	Detecting	False alarm
8000 – 12000	5 from 5	2 from 60
500– 600	0 from 5	
400 – 500	0 from 5	

4.3. Internet SG (SG-I) [5].

This type of SG is based on embedding into different Internet protocols like TCP/IP.



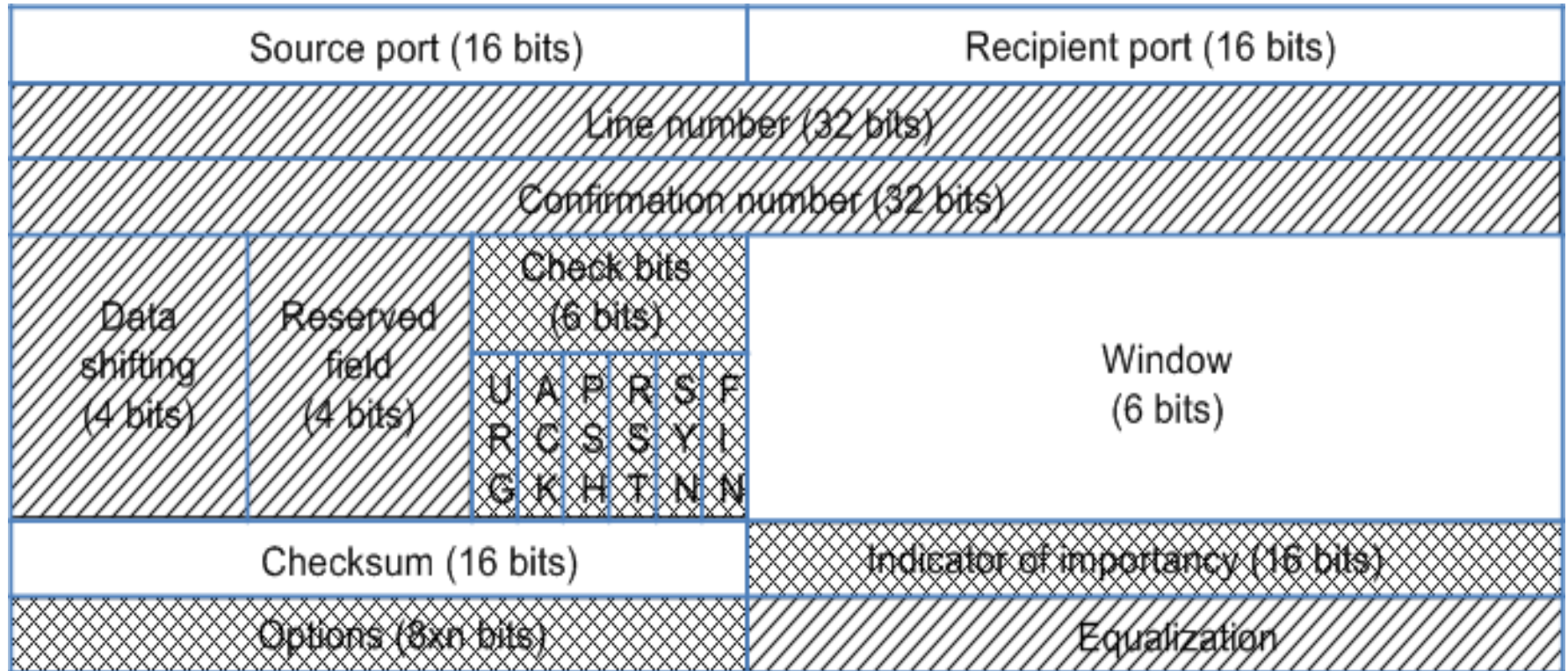
Embedding is possible on all OSI levels .

OSI Internet architecture

Methods of embedding on different levels.

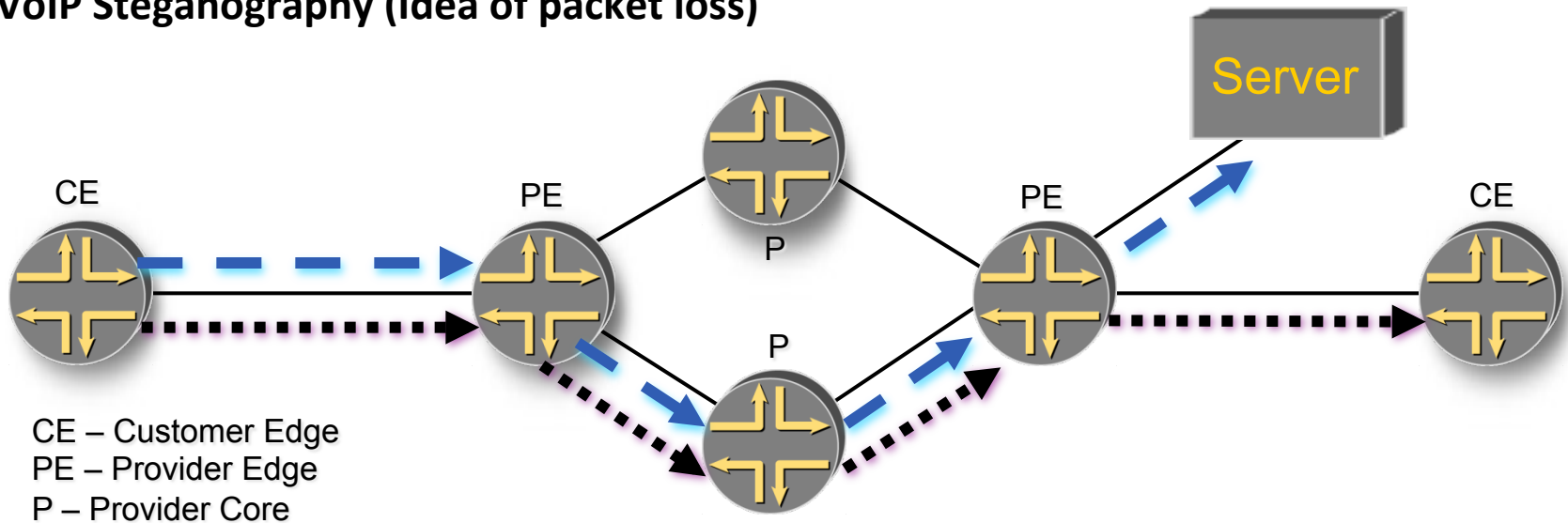
Application	The use of conventional steganography
Presentation	Embedding into the fields of system messages
Session	Monitoring of remote sites
Transport	Embedding into unused heads of TCP protocols
Network	Embedding into free fields of IP packets
Data-link	Embedding into heads of frames , the use of CRC information
Physical	The use of conflict situations: “0” – to send packet just after delay , “1” – to send packet just after conflict

Format of TCP head.



The fields in which is allowed to embed information unconditionally are shown by hatching whereas the field where embedding is allowed under some conditions are shown by double hatching .

VoIP Steganography (Idea of packet loss)



The main idea is to organize simulated “packet loss” between PE-routers:

1. Customer A started RTP-session to Customer B
2. Provider analyses the codec and calculate how many packets from this session may be lost without degradation of speech:
 - G.711 codec - <10% packet loss
 - G.729 codec - <1% packet loss
3. Provider selects a packet , inserts needed information and send it to other side of provider network or other provider
4. Opposite PE needs to know the number of changed packet and to be able to route it to needed interface
5. Customer sees it as normal packet loss in provider network