DETERMINING THE AMOUNT OF HIDDEN INFORMATION IN JPEG IMAGES

Hristo T. Terzev, Zhaneta N. Savova

Department of Computer Systems and Technologies, Facylty of Artillery, Air Deffence and Communication and Information Systems, National Millitary University, Shumen, Bulgaria, zh.savova@mail.bg

***Abstract**: The application of visual, structural, statistical and analytical steganalysis detects the presence of hidden information in images, without being able to determine its amount. Based on the advantages of the JPEG format over BMP and PNG formats and the fact that the JPEG format is set by default for most digital cameras, this paper answers the question how RS steganalysis can be applied to determine the amount of hidden information in JPEG images. A methodology for examining the amount of hidden information in JPEG images is proposed and the results of the conducted steganalysis are presented.*

***Keywords**: Steganalysis, RS Analysis, JPEG images, LSB Steganography*

ОПРЕДЕЛЯНЕ НА КОЛИЧЕСТВОТО СКРИТА ИНФОРМАЦИЯ В JPEG ИЗОБРАЖЕНИЯ

Христо Т. Терзиев, Жанета Н. Савова

Въведение

Внедряването на цифровите технологии и широкото приложение на информационните и комуникационни технологии [1], [2] в почти всички области на човешката дейност са фактори, които направиха цифровите изображения едни от най-популярните предавани файлове в глобалната световна мрежа Интернет. Поради големия размер на излишък в цифровото им представяне, промяната на който е визуално незабележима, и благодарение на спецификата на човешката зрителна система, скриването на конфиденциални съобщения в тях е един от най-предпочитаните варианти гарантиращи секретност в комуникациите. Тези техники, известни като стеганографски методи, са подход в областта на информационната сигурност за осигуряване на поверителност на информация при скриването ѝ в обект контейнер. Те осигуряват неоткриваемост на факта, че комуникационният канал предава скрита информация за неоторизирани прехващачи, което намалява вероятността за откриване на скритата информация.

Цифровите изображения представят дигитално двумерна информация, която посредством алгоритъм се преобразува в специален двоичен код на стандартизиран файлов формат. Стеганографско вграждане на скрити данни във файлови формати като BMP и PNG не е удачно поради големия им размер при транспортиране. Стандарт в цифровата фотография и онлайн споделянето на изображения е JPEG форматът, осигуряващ възможности за съхраняване с настройваща се загуба на качество при внимателно балансиране на размера на файла с постигането на висока

разделителна способност, висока дълбочина и яснота на цветовете. Няколко милиарда JPEG изображения се създават всеки ден [3]. Според статистиките от 01.01.2012 до 19.09.2020 г. в [11] на тенденциите за използване на графични файлови формати за уебсайтове, JPEG е един от най-разпространените формати.

От друга страна методите на стеганализ са приложими в случаите, когато се цели да се засече и разпознае скрита в изображенията информация. В резултат могат да се осуетят действията на вътрешни лица, които използват стеганографията като инструмент за предаване на чувствителна и класифицирана информация. Методите са ефективни и за прихващане на съобщения за организиране на престъпна дейност от терористи и трафиканти на наркотици. По своята същност прилагането на визуален, структурен, статистически и аналитичен стеганализ подават сигнал за това, че в контейнера може би е скрита някаква информация, без да могат да определят нейното количество. По своята същност предложеният през 2001 г. RS сеганализ [7], [8] може да определи дължината на скритото съобщение, вградено посредством метода на най-младшия бит в LSB в BMP изображения.

Изхождайки от предимствата на JPEG формата [6] и [12] и факта, че форматът е настроен по подразбиране за повечето цифрови фотоапарати, в настоящата статия се отговаря на въпроса как RS сеганализът може да се приложи за определяне на количеството скрита информация в JPEG изображения. Статията е организирана по следния начин. Първо се анализира JPEG файловия формат. След това накратко се представя същността на RS стеганализа. В третата част се предлага методика за изследване на количеството скрита информация в JPEG изображения и се представят резултатите от проведения анализ. Статията завършва с обобщаване на резултатите.

Анализ на JPEG файлови формати

JPEG изображенията са дискретни. Всеки от пикселите им съдържа трите компонента на RGB цветовия модел и при компресиране се трансформира в YCbCr цветово поле, при което CbCr са цветовите канали, а Y е яркостта. Връзката между разделителната способност на JPEG контейнера, дълбочината на цвета в RGB модела и максималния брой символи на скритото съобщение, което може да бъде вградено е:

$$\text{Брой символи} = \frac{\text{брой пиксели по хоризонтала x брой пиксели по вертикала } x\ 3}{8} \tag{1.1}$$

Вграждането на скритите данни в DCT коефициентите на JPEG изображенията се извършва на два етапа. Първият етап E_1 не е обратим и изчислява с даденото съотношение на компресия квантованите DCT коефициенти.

$$\boldsymbol{E_1}\textit{: (JPEG контейнер, коефициент на компресия)} \rightarrow \textit{DCT коефициенти} \tag{1.2}$$

Единствената част от алгоритъма на компресиране, при която може да се вгради секретна информация и в последствие изцяло да се възстанови е обратимия етап E_2, преобразуващ в двоичен код b_i квантованите DCT коефициенти.

$$\boldsymbol{E_2}\textit{: (DCT коефициенти)} \rightarrow b_i,\ b_i \in Z_2^* \tag{1.3}$$

Битовете на скритите данни са произволно разпространени по цялото JPEG изображение, според пермутация, зависима от потребителската парола, която ги вмъква в уникален ред в DCT коефициентите. Невъзможно е при χ^2 тест, без наличието на оригинален ключ, да се определи в кой от DCT коефициентите е вграден таен бит от скритото послание.

Възможности на RS стеганализ за разпознаване на стегоизображения

Оригинален алгоритъм за статистически стегоанализ е Regular-Singular (RS) анализа [7], който е предложен за пръв път през 2001 г. от екип от учени, ръководен от Джесика Фридрих. RS анализът е математически формулиран, изследван и ефективността му е проверена в [4], [5], [9],

[10] и [13]. Този метод търси скрити зависимости между елементите на контейнера. За целта цялото изображение се разделя на групи от малък брой четни n пиксели (например 2 или 4), които са разположени един до друг хоризонтално. Групата от пиксели $G = (x_1, x_2, ..., x_n)$ се състои от пикселите $x_1, x_2, ..., x_n$ със стойности от 0 до 255, съответстващи на 8-битов цветови канал. За групата G се дефинира функцията за регулярност или „гладкост“ $f(G)$, която определя дисперсията на пикселите в групата, т.е. сумата от разликите в стойностите на съседните пиксели. Смяната на младшите битове на пикселните стойности с битовете на скритото съобщение добавя шум към изображението. При нарастването на нивото на шум в пикселите на групата, стойността на $f(G)$ нараства.

$$f\ x_1, x_2, \ldots., x_n \ = \ \sum_{i=1}^{n=1} x_{i+1} - x_i \ , \tag{3.1}$$

Функцията $F(x)$ е обратна, като $F(f(x)) = x$. Дефинират се две пермутации. Първата F_1 съответства на инверсията на най-младшия бит на пиксела и F_{-1}, която е инверсия с пренос към старшия бит (добавян е един):

$$F_1\colon 0 \leftrightarrow 1, 2 \leftrightarrow 3, \ldots, 254 \leftrightarrow 255$$
$$F_{-1}\colon 255 \leftrightarrow 0, 1 \leftrightarrow 2, 3 \leftrightarrow 4, \ldots, 253 \leftrightarrow 254, 255 \leftrightarrow 0 \tag{3.2}$$

С помощта на обратима функция за дадена група пиксели G се дефинират следните видове групи:

- Редовни групи (R-групи): $G \in R \Leftrightarrow f(F(G)) > f(G)$
- Единични групи (S-групи): $G \in S \Leftrightarrow f(F(G)) < f(G)$
- Неизползвани групи: $G \in U \Leftrightarrow f(F(G)) = f(G)$,

където обратимата функция $F(G)$ се прилага към всички пикселни стойности на $G = (x_1, x_2, ..., x_n)$.

Характерно за графичните изображения без съдържание на секретни данни е по-големия брой редовни R групи в сравнение с единичните S групи. Може да се твърди за наличие на скрита информация, чрез LSB стеганография в най-младшите битове на пикселите, когато за стойностите на R_M и S_M в изследваното изображение се наблюдава зависимостта:

$$R_M \cong S_M, \tag{3.3}$$

където R_M е относителната стойност на броя на редовните R групи с маска M спрямо общия брой в групата, в проценти, а S_M е съответно относителната стойност на броя на единичните S групи с маска M спрямо общия брой в групата, в проценти.

Средноаритметичната стойност на величината x по абсцисата на координатите на пресечната точка на R_M и S_M позволява да се определи големината на вграденото съобщение p чрез формулата:

$$p = \frac{x}{x-0,5}, \tag{3.4}$$

По този начин изходната стойност от резултата от RS анализа показва не само наличието на конфиденциална информация, но и нейния размер. Методът е ефективен при последователно и разпръснато LSB вграждане.

Определяне на количеството скрита информация след JPEG трансформации

За целите на реализираното изследване са използвани контейнери в JPEG файлов формат с различна резолюция и размер. Информацията в тях е скрита с програмния продукт Virtual Steganographic Laboratory. Трансформациите са извършени със софтуерния пакет за цифрова обработка на изображения PhotoScape. Получените стегоизображения са анализирани с StegSecret.

Методиката на изследване за реализиране на експериментите е следната:

1. В контейнерите, избрани JPEG изображения, последователно се вгражда секретно съобщение с капацитет от 10 % до 100 % от големината на контейнера, със стъпка 10 %.

2. Получените JPEG стегоизображения се преобразуват в BMP файлов формат.
3. Получените BMP изображения се тестват с RS стеганализ за определяне на количеството скрита информация в тях.
4. Контролен тест за сравнение с RS стеганализ се реализира на всеки избран контейнер.

Целта на изследването е да се провери до каква степен може да се определи количеството скрита информация в JPEG изображения с помощта на RS стеганализ преди вграждане на информация и при различни нива на вградена информация. Едновременно с това изследването цели и да провери устойчивостта на JPEG файловия формат срещу трансформации, при които се губи част от мултимедийното му съдържание.

Изследването е реализирано със 100 различни JPEG изображения, като в таблица 1 са представени резултатите за 3 от изображенията при реален размер на вградената информация от 10 % до 100 %.

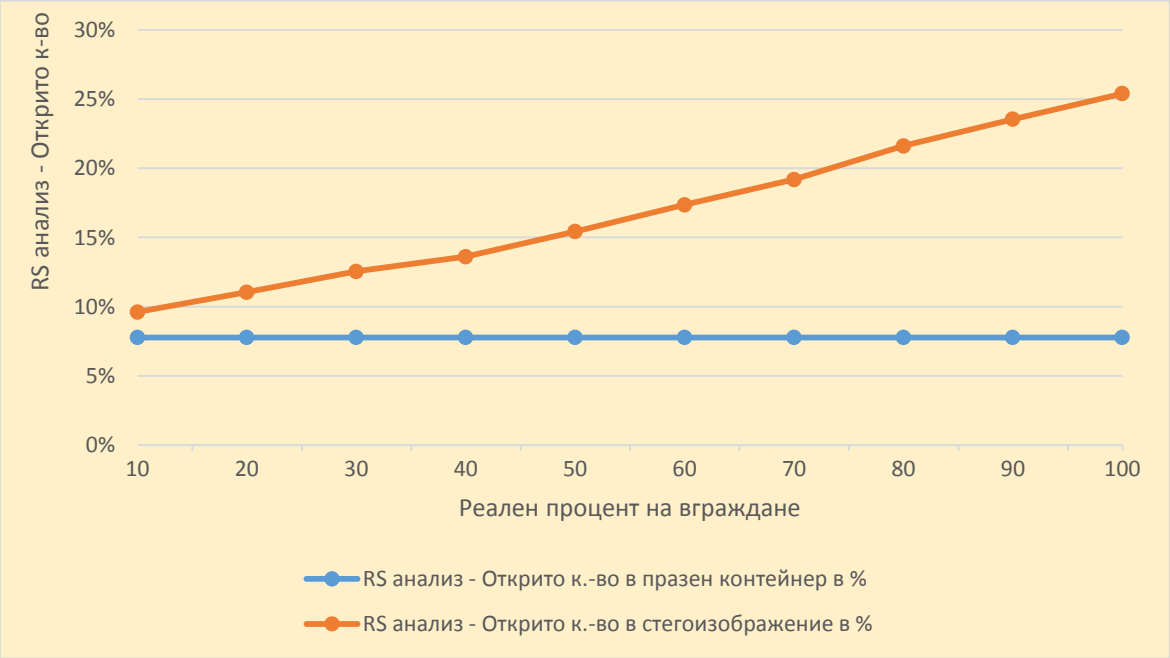
Табл. 1. *Изследване на скрита информация в JPEG контейнери*

№	Изображение	Празен контейнер		Реален размер на вграждане	Стегоизображение			
		RS анализ Открито количество	Средно отклонение спрямо реалния размер + / -		RS анализ Открито количество	RS анализ Средно открито количество	Отклонение спрямо реалното вграждане + / -	Средно отклонение спрямо реалното вграждане + / -
1	SAM_4500.jpeg	7,77%	3,23%	10%	9,62%	5,69%	-0,38%	-4,31%
2	SAM_4597.jpeg	1,82%			6,20%		-3,80%	
3	SAM_4644.jpeg	0,09%			1,26%		-8,74%	
4	SAM_4500.jpeg	7,77%	3,23%	20%	11,04%	7,88%	-8,96%	-12,12%
5	SAM_4597.jpeg	1,82%			9,92%		-10,08%	
6	SAM_4644.jpeg	0,09%			2,69%		-17,31%	
7	SAM_4500.jpeg	7,77%	3,23%	30%	12,55%	9,69%	-17,45%	-20,31%
8	SAM_4597.jpeg	1,82%			12,20%		-17,80%	
9	SAM_4644.jpeg	0,09%			4,31%		-25,69%	
10	SAM_4500.jpeg	7,77%	3,23%	40%	13,62%	11,27%	-26,38%	-28,73%
11	SAM_4597.jpeg	1,82%			14,05%		-25,95%	
12	SAM_4644.jpeg	0,09%			6,13%		-33,87%	
13	SAM_4500.jpeg	7,77%	3,23%	50%	15,43%	12,90%	-34,57%	-37,10%
14	SAM_4597.jpeg	1,82%			15,58%		-34,42%	
15	SAM_4644.jpeg	0,09%			7,70%		-42,30%	
16	SAM_4500.jpeg	7,77%	3,23%	60%	17,36%	14,72%	-42,64%	-45,28%
17	SAM_4597.jpeg	1,82%			17,35%		-42,65%	
18	SAM_4644.jpeg	0,09%			9,44%		-50,56%	
19	SAM_4500.jpeg	7,77%	3,23%	70%	19,19%	16,48%	-50,81%	-53,52%
20	SAM_4597.jpeg	1,82%			18,92%		-51,08%	
21	SAM_4644.jpeg	0,09%			11,35%		-58,65%	
22	SAM_4500.jpeg	7,77%	3,23%	80%	21,62%	18,48%	-58,38%	-61,52%
23	SAM_4597.jpeg	1,82%			20,27%		-59,73%	
24	SAM_4644.jpeg	0,09%			13,54%		-66,46%	
25	SAM_4500.jpeg	7,77%	3,23%	90%	23,54%	20,75%	-66,46%	-69,25%
26	SAM_4597.jpeg	1,82%			22,03%		-67,97%	

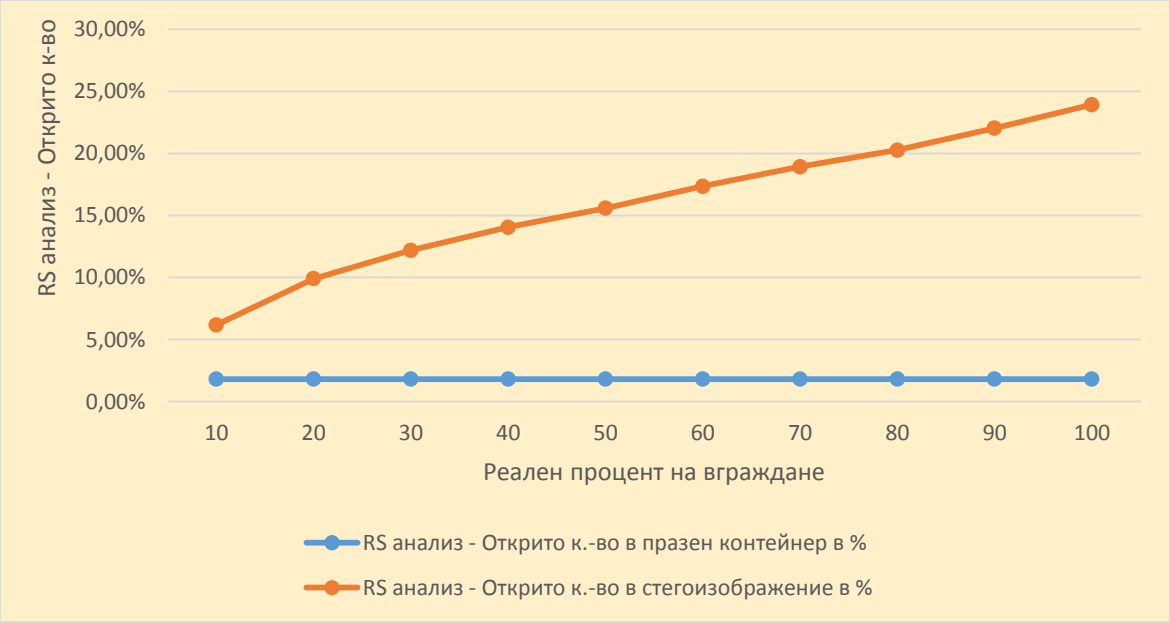
№	Изображение	Празен контейнер		Реален размер на вграждане	Стегоизображение			
		RS анализ Открито количество	Средно отклонение спрямо реалния размер + / -		RS анализ Открито количество	RS анализ Средно открито количество	Отклонение спрямо реалното вграждане + / -	Средно отклонение спрямо реалното вграждане + / -
27	SAM_4644.jpeg	0,09%			16,69%		-73,31%	
28	SAM_4500.jpeg	7,77%			25,40%		-74,60%	
29	SAM_4597.jpeg	1,82%	3,23%	100%	23,93%	23,78%	-76,07%	-76,22%
30	SAM_4644.jpeg	0,09%			22,00%		-78,00%	
Среден резултат				55,00%	-	14,16%	-	-40,84%

Диаграмите на фигури 1, 2 и 3 отразяват визуално резултатите за трите JPEG изображения от таблица 1. Може да се направят следните изводи:

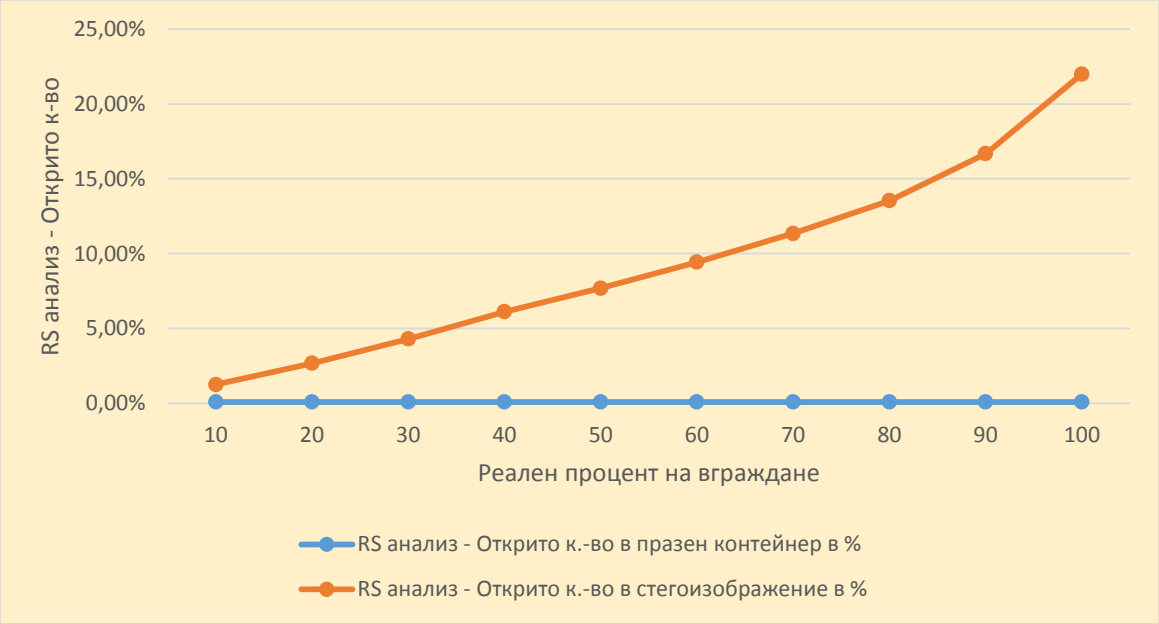
1. Извършеният RS анализ върху празните JPEG контейнери открива процент вградена информация средно около 3,23 %, която се дължи на преобразуването на JPEG формата в BMP.
2. Извършеният RS анализ може да определя количеството вградена информация в JPEG изображения като се запазва тенденцията за 40,84 % по-малко количество от реалното.



Фиг. 1. *RS анализ на SAM_4500.jpeg контейнер*



Фиг. 2. *RS анализ на SAM_ SAM_4597.jpeg контейнер*



Фиг. 3. *RS анализ на SAM_4644.jpeg контейнер*

Заключение

Методите на визуален, структурен, статистически и аналитичен стеганализ, засичащи скритата информация в изображения, са ефективни за осуетяване действията на вътрешни лица, използващи стеганографията за предаване на чувствителна и класифицирана информация, и за прихващане на съобщения за организиране на престъпни дейности. Предложената в статията методика за определяне на количеството скрита информация в JPEG изображения се основава на предложения през 2001 г. RS стеганализ и преобразуване на JPEG стегоизображението в BMP формат. Експерименталните резултати от направените изследвания показват, че RS стеганализът

открива наличието на количество скрита информация в JPEG изображения, което е с 40.84% по-малко от действителното вградено.

References

1. Богданов, Р. А. (2012) Когнитивното радио – следващ етап в развитието на радиокомуникациите. *CIO, бр. 9, 2012*, ISSN 1312-5605, Retrieved from http://cio.bg/4890_kognitivnoto_radio__sledvasht_etap_v_razvitieto_na_radiokomunikaciite
2. Богданов, Р. А. (2011) Развитие на стандартите за безжични компютърни комуникации. *Сборник научни трудове на научна конференция на Факултет „А, ПВО и КИС", 13-15 ноември 2010, част 1 „Комуникационни и информационни системи"*, Шумен, Химера, 2011, стр. 175-183, ISSN 1313-7433
3. Baraniuk, C. (15 October 2015). Copy protections could come to JPEGs. BBC News. BBC.
4. Cancelli, G., G. Doerr, I. Cox, & M. Barni (2008) Detection of ±1 LSB steganography based on the amplitude of histogram local extrema. *15th IEEE International Conference on Image Processing (ICIP)*, San Diego, pp. 1288 - 1291.
5. Cox, I. J., M. Miller, J. Bloom, J. Fridrich, & T. Kalker. (2008) *Digital Watermarking and Steganography, 2nd ed.*: Morgan Kaufmann Publishers.
6. Eltyeb E, &A. Elgabar. (2013) Comparison of LSB Steganography in BMP and JPEG Images. *International Journal of Soft Computing and Engineering (IJSCE)*, ISSN: 2231-2307, Volume-3, Issue-5.
7. Fridrich, J., Goljan, M., & Du, R. (2001). Detecting LSB steganography in color, and gray-scale images. *IEEE multimedia*, *8*(4), 22-28.
8. Fridrich, J., Goljan, M., Hogea, D., & Soukal, D. (2003). Quantitative steganalysis of digital images: estimating the secret message length. *Multimedia systems*, *9*(3), 288-302.
9. Fridrich, J. & M. Long. (2000) Steganalysis of LSB Encoding in Color Images. *IEEE Int. Conf. on Multimedia and Expo*, 2000, pp. 1279-1282.
10. Geetha, S., S. Sindhu, R. Renganathan, P. Raman, & N. Kamraj. (2008) StegoHunter: Steganalysis of LSB Embedded Images based on Stego-Sensitive Threshold Close Color Pair Signature. *Sixth Indian Conference on Computer Vision, Graphics & Image Processing, ICVGIP '08*, pp. 281-288.
11. Q-Success. (29 September 2020) Historical yearly trends in the usage statistics of image file formats for websites. Retrieved from https://w3techs.com/technologies/history_overview/image_format/all/y
12. Sinha, B. (2015) Comparison of PNG and JPEG format for LSB Steganography, *International Journal of Science and Research, vol. 4, no. 4,5*, pp. 198-201.
13. Westfeld A. & A. Pfitzmann. (1999) Attacks on Steganographic Systems. *Proc. Of 3rd Information Hiding Workshop*, Dresden, Germany, pp. 61 -75.