Magdalena Pejas

Techniki ukrywania informacji w danych cyfrowych i narzędzia je wykrywające

Już od starożytności w celach militarnych stosowano steganografię, czyli technikę ukrywania wiadomości w powszechnie dostępnych przedmiotach. Używano do tego celu na przykład atramentu sympatycznego, wosku czy tatuaży. W czasie drugiej wojny światowej do ukrywania niejawnych komunikatów stosowano mikro-kropki.

Słowo „steganografia" pochodzi z języka greckiego i oznacza *ukryte pismo.* Innymi słowy jest to sztuka ukrywania informacji w przedmiotach lub wiadomościach powszechnie dostępnych.[1)] W odniesieniu do wiadomości i danych cyfrowych będzie to sztuka zapisywania niejawnych danych cyfrowych (tekstów, plików) w innych jawnych plikach tak, aby nie było to widoczne dla nieuprawnionego użytkownika. Tylko wyznaczony odbiorca znający sposób zakodowania danych może je odzyskać i odczytać.

Obecnie rozwijające się technologie internetowe (*e-poczta, e-banki, e-zakupy, e-książki, etc.*) umożliwiają nieomal nieograniczone możliwości przesyłania danych cyfrowych. Sprzyja to wyciekom informacji z firm, niekontrolowanej wymianie informacji pomiędzy podejrzanymi organizacjami i pomocy w przestępstwach zorganizowanych. Każdą informację można praktycznie ukryć wszędzie, przemycić omijając zaawansowane zabezpieczenia.

W Internecie jest dostępnych wiele instrumentów steganograficznych. Gros tych programów umożliwia ukrywanie plików w dowolnym formacie (np.: tekst zwykły pisany z klawiatury, dokument edytora Microsoft Word, arkusz programu Microsoft Excel), w dużych plikach graficznych i w multimediach. Należy tu podkreślić, że ciągle wzrasta szybkość przesyłania, wielkość pamięci i dostępność danych. Ogromne zróżnicowanie standardów i formatów danych sprzyja technikom ukrywania bądź przemycania wiadomości a nawet całych baz danych.

Poniżej zaprezentowano proste techniki ukrywania wiadomości i danych, w systemie Windows i pakiecie MS Office. W dalszej części przedstawiono przykłady internetowych narzędzi służących do wykrywania przekazów steganograficznych oraz własne indywidualne metody wykrywania śladów użycia wybranych prostych programów steganograficznych.

Funkcja steganograficzna jest zwykłym programem komputerowym, który daje możliwość ukrywania danych, np. w obrazku czy zdjęciu cyfrowym. Następnie trzeba podać ścieżkę dostępu do tego obrazka. W wyniku przetworzenia pliku po ukryciu w nim danych wygląda tak samo jak oryginał. Następnie zdjęcie nośne zostaje wysłane do odbiorcy, który ma taki sam program steganograficzny.

Przykładem może być gra komputerowa, która zawiera dużo animacji, dźwięku i obrazów, więc istnieje możliwość ukrycia tam stosunkowo dużej ilości danych. W takim przypadku:

1. tworzy się program steganograficzny, którego algorytm ukrywania danych jest taki, że prawdopodobieństwo odczytania ukrytych informacji przez przypadkowego użytkownika jest praktycznie niemożliwe,

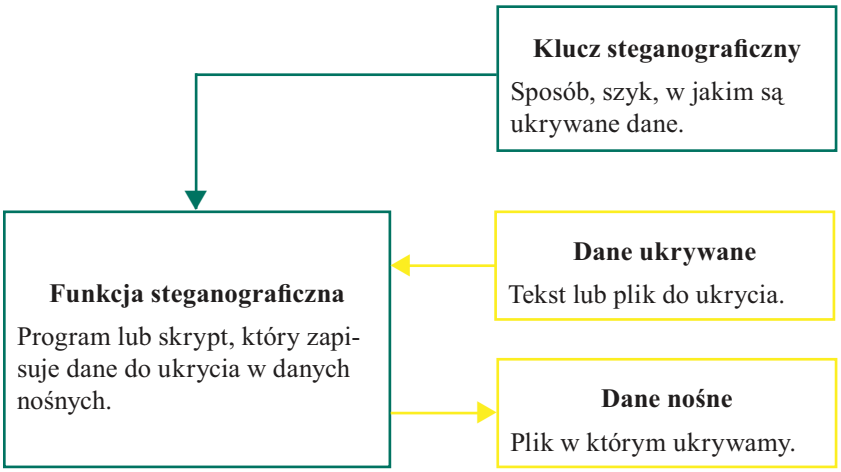
[1)] http://pl.wikipedia.org/wiki/Steganografia

2. w partii egzemplarzy gry zostaje ukryty tekst i rysunki,
3. nadawca udostępnia grę do sprzedaży,
4. tylko wybrane osoby wiedzą, jakim algorytmem i z jakim hasłem ukryto dane.

Istnieją tez programy do podejmowania prób wykrywania ukrytych przekazów, ale są one mało skuteczne.

Pojęcie	Opis
Dane do ukrycia	Dowolny tekst lub plik dowolnego formatu, który mamy ukryć w innym większym pliku.
Dane nośne	Dowolny plik o odpowiedniej wielkości, w którym mamy ukryć dane bez zmiany wyglądu i sposobu działania tego pliku.
Funkcja steganograficzna	Program lub skrypt zapisujący dane do ukrycia w danych nośnych.
Algorytm steganograficzny	Sposób działania programu ukrywającego dane.
Klucz steganograficzny	Sposób rozmieszczania danych do ukrycia w pliku nośnym.
Pojemność danych nośnych	Maksymalna ilość danych do ukrycia jaką można „schować" w pliku nośnym bez zauważalnej zmiany jego wyglądu, struktury i działania.

Tab. 1. Podstawowe pojęcia związane z ukrywaniem wiadomości lub danych cyfrowych



Rys. 1. Klasyczny system steganograficzny

Steganografia w Polsce i na świecie

Na podstawie wiadomości z Internetu można wysnuć wniosek, że najwięcej badań na temat technik steganograficznych prowadzi się na uczelniach zagranicznych. W Polsce temat ten dopiero w ostatnich latach zaczął się pojawiać w postaci prac magisterskich i doktorskich, między innymi na Politechnice Warszawskiej. Światowe publikacje na ten temat pochodzą z krajów takich jak Niemcy (*Rostock University*), Wielka Brytania (*University of St Andrews*) czy Irlandia (*University of Ulster*), jak również Stany Zjednoczone Ameryki (*Harvard University*, *University of Kalifornia, University of Michigan*).

Techniki ukrywania informacji

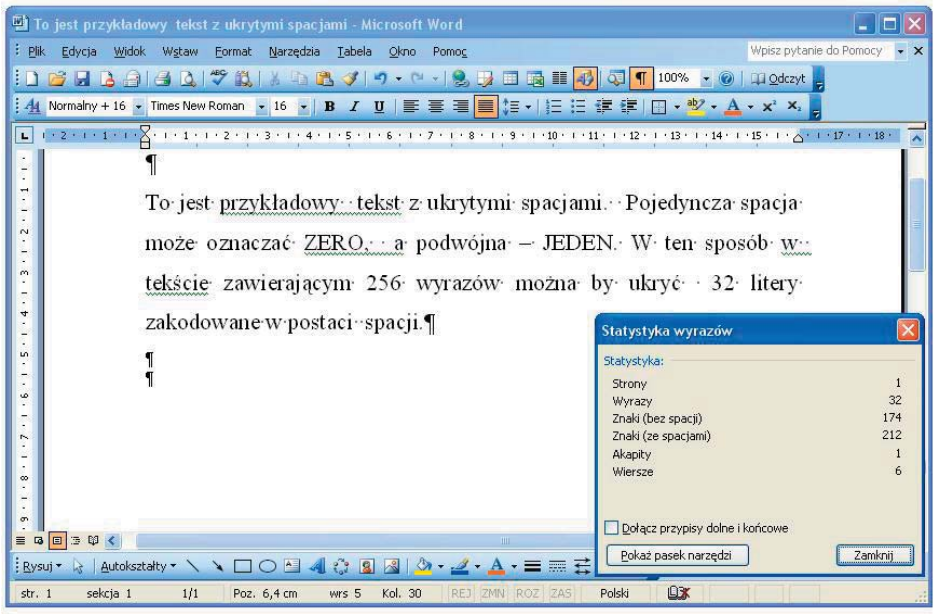
Istnieje wielka dowolność formatów plików, które mogą posłużyć za dane nośne. Są to między innymi:

- pliki tekstowe (txt, html, xml),
- pliki MS Office (Word, PowerPoint, Excel, Access),
- grafika (bmp, jpg, gif),
- pliki dźwiękowe (wav, mp3),
- filmy wideo (avi),
- dane DVD (mpg2),
- dane komunikacji sieciowej (pakiety sieciowe),
- inne (historia transakcji finansowych, bilingi telefoniczne).

Dane z tych kategorii mogą być oczywiście także ukrywane w innych danych o odpowiednim rozmiarze. Na uwagę zasługuje kategoria „inne". Do ukrytej komunikacji można używać wartości przelewów bankowych, czy też czas trwania rozmowy telefonicznej. Urywane sygnały telefoniczne można porównać do kodowania pojedynczych informacji typu „TAK" albo „NIE". Teoretycznie przy ich użyciu można nadać dowolne serie znaków.

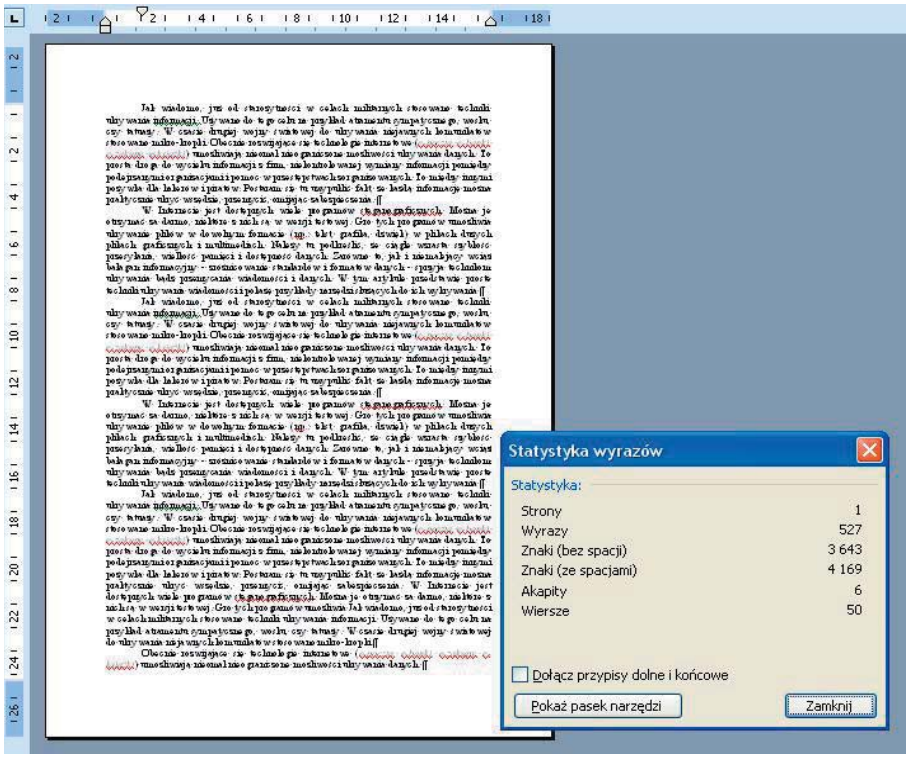
Poniżej przedstawiono trzy przykłady ukrywania tekstu w tekście:

- podwójne spacje pomiędzy wyrazami w dokumencie Microsoft Word;
- spacje w znacznikach w stronie html;
- dowolne dane w pliku graficznym.



Rys. 2. Podwójne spacje w tekście w Microsoft Word

Przeciętna wypełniona tekstem strona (czcionka *Times New Roman*, 12 pkt) zawiera ponad 500 wyrazów.

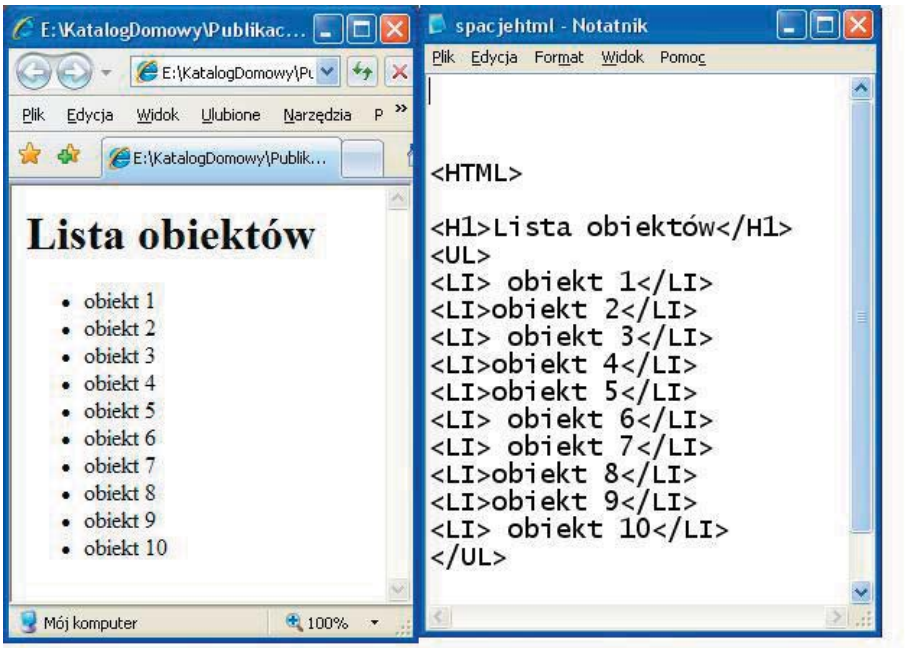


Rys. 3. Statystyka wyrazów na jednej stronie w dokumencie w edytorze Microsoft Word

Oznacza to, że można w nim ukryć ponad 500 zer i jedynek w postaci spacji. Ponieważ każda litera alfabetu, których jest 28, może być zakodowana serią pięciu zer lub jedynek, na jednej stronie dokumentu można ukryć około 100 liter, co odpowiada ponad jednej linijce ukrytego tekstu. Nie jest to dużo, ale są jeszcze bardziej wydajne sposoby ukrywania niewidzialnych tekstów w dokumentach. Przykładem są strony internetowe, które w notatniku wyświetlane są w postaci specjalnych znaczników języka HTML czytelnych dla przeglądarek internetowych.

Ukrywanie białych znaków w stronach internetowych

Rysunek 4 przedstawia, w jaki sposób można ukryć dodatkowe spacje w pliku internetowym w znacznikach określających elementy listy. Przy przeglądaniu tego pliku w przeglądarce internetowej są one niewidoczne, natomiast w Notatniku lub w Microsoft Word można je zobaczyć. W tym przypadku można ukryć nawet długie ciągi tzw. „białych znaków", czyli takich, których nie widać w przeglądarce, ale są one czytelne dla programów komputerowych, np. na seriach czterech białych znaków, które mogą występować w trzech postaciach (spacja, tabulator, enter) można zapisać aż 81 różnych znaków w ten sposób, że każdej kombinacji białych znaków zostaje przypisana inna litera alfabetu.



Rys. 4. Przykładowa strona HTML i widok źródłowy z ukrytymi spacjami

W pliku strony internetowej można ukryć dowolnie wiele „białych znaków", które nie są widoczne przy przeglądaniu go w przeglądarce internetowej, natomiast łatwo może je wychwycić odpowiedni program steganograficzny.

Ukrywanie danych w plikach graficznych

Najodpowiedniejszym przykładem jest obrazek, który w oryginale ma rozmiar 600 na 800 pikseli. Każdy piksel ma trzy barwy, w odcieniach od 0 do 255. Modyfikując te barwy tyko o 1, można zakodować serię zer i jedynek w ilości 3xwysokośćxszerokość obrazka (w pikselach), co w przypadku tego obrazka daje 3*600*800, co równa się 1.440.000 bitów (podstawowych jednostek informacji). Przy założeniu, że 32 znaki (alfabet plus wybrane znaki specjalne) zakodujemy na 5-ciu bitach (zero albo jeden), to zajmując cały obszar obrazka, zakodujemy 288 tysięcy znaków bez zauważalnego naruszenia jego wyglądu. Jeśli przyjąć, że w przeciętnym dokumencie programu Microsoft Word na jednej stronie znajduje się około 4000 liter, to w pokazanym niżej obrazku można „schować" około 72 stron dokumentu (!).

Przeciętna pojemność programów do ukrywania danych w obrazkach wynosi około 15%. Tak więc jeśli informacja zostanie ukryta w każdym pikselu barwy na najmniej znaczącym bicie dla każdej składowej barwy, pojemność danych nośnych wyniesie wtedy: $^1/_8$=12,5%, zaś jeśli na dwóch, wyniesie ona ¼, czyli 25%. Gdyby zamiast tekstu ukryć plik innego typu, to jego „bezpieczna" wielkość (dla wspomnianego wyżej pliku graficznego) wynosiłaby 180 kB (przy wykorzystaniu $^1/_8$ wielkości pliku nośnego).

Należy jeszcze raz podkreślić, że nie stanowi to żadnej różnicy ani dla oka ludzkiego, ani dla większości programów wykrywających anomalia w obrazie pod warunkiem, że nie przekroczy się dopuszczalnej pojemności graficznych danych nośnych, które zależą od rozmycia, kontrastów i gładkości obrazów.

Programy steganograficzne w sieci

W sieci Internet jest dostępnych wiele programów do ukrywania danych, niektóre są bezpłatne i bardzo proste w użyciu. W tabeli 2 zebrano kilka przykładowych programów steganograficznych wyszukanych w Internecie.

Nazwa programu
Courier 1.0,
Image Hide,
JPHS for Windows 0.5,
Hide In Picture 2.0,
Steg Hide 0.5.1,
Steganography 1.6.5.
WbStego 4.2

Tab. 2. Wybrane programy steganograficzne dostępne w Internecie

Większość z wymienionych programów służy do ukrywania danych w plikach graficznych (bmp i jpg). Rzadziej spotyka się programy, które ukrywają dane w plikach muzycznych (wav, mp3).

W tabeli 3 zebrano kilka najważniejszych adresów internetowych, gdzie można znaleźć obszerne informacje nie tylko wyjaśniające pojęcia związane ze steganografią, ale zawierające również odnośniki do osób (*Neil Johnson, Niels Provos, Fabien Petitcolas*) oraz instytucji (*Johnson & Johnson Technology Consultants LLC*) zajmujących się tą dziedziną a także do programów steganograficznych i stegano-analitycznych.

Nr	**Adresy internetowe**
1	http://www.en.wikipedia.org/wiki/Steganography
2	http://www.jjtc.com/Steganography
3	http://www.jjtc.com/stegdoc
4	http://www.petitcolas.net/fabien/steganography
5	http://www.umich.edu/u/provos/stego
6	http://niels.xtdnet.nl/stego
7	http://www.outguess.org
8	http://www.tech-faq.com/steganography.shtml
9	http://stegoarchive.com
10	http://www.forensics.nl/steganography

Tab. 3. Odnośniki poświęcone steganografii i bezpieczeństwu informacji

W Internecie znajdują się informacje o światowych konferencjach tematycznie związanych ze steganografią i cyfrowym znakowaniem wodnym. Przykładowe wyszukane w sieci wystąpienia nazw różnych konferencji zebrano w tabeli 4.

Nr	Nazwy międzynarodowych konferencji
1	International Information Hiding Conference
2	International Conference on Intelligent Information Hiding
3	International Information Security Conference
4	International Conference on Image Processing
5	IEEE Conference on Multimedia

Tab. 4. Światowe konferencje

Rola systemu Windows w ukrywaniu wiadomości

Warto zauważyć fakt, że do ukrywania danych świetnie nadają się pliki utworzone za pomocą pakietu MS Office. Są to pliki pocztowe, tekstowe, prezentacje multimedialne, arkusze kalkulacyjne, a nawet bazy danych. Dostęp do tych plików można zautomatyzować za pomocą specjalnych makr czy skryptów uruchamianych w systemie Windows i napisanych w języku Visual Basic. Nie wymagają one instalacji dodatkowych programów. Dla kogoś, kto zna ten język programowania chociaż w stopniu podstawowym i orientuje się w makrach, zupełnie naturalnym stanie się zupełnie swobodny i szybki dostęp do następujących czynności:

- znajdowanie plików według wybranych kryteriów (typ, obecność danego tekstu),
- odczytanie informacji o danym pliku (rozmiar, data utworzenia),
- przeczytanie nazw i parametrów wszystkich tabel, wykresów, slajdów, rysunków w danym pliku,
- wczytanie treści tekstowej pliku i wyszukanie w nim istotnych informacji,
- modyfikacja parametrów obiektów w plikach (ramki tekstu, tabele, rysunki),
- zmodyfikowanie zawartości pliku w dowolny sposób (akapity, tabele, wykresy, rysunki, zdjęcia),
- zmiana atrybutów pliku (data ostatniej modyfikacji).

Wspomniane skrypty umożliwiają również automatyczne tworzenie lub modyfikację plików stron internetowych. Jeśli w danym skrypcie umieszczony jest algorytm steganograficzny, umożliwia to swobodne wyszukiwanie informacji, odczytywanie treści i modyfikację zawartości oraz parametrów plików. Przykładem może być specjalnie napisany skrypt do przeczytania wszystkich adresatów z programu Microsoft Outlook a następnie do umieszczenia ich listy w postaci ukrytej w prezentacji Microsoft PowerPoint, poprzez minimalną manipulację rozmiarami i położeniem obiektów na slajdach.

Warto podkreślić, że skrypty działające pod MS Office umożliwiają uzyskanie treści z plików Word, Excel, etc., bez czytania ich w całości, zatem automatycznie można zaoszczędzić na wielkości ukrywanych danych, bo nie muszą one zawierać informacji o budowie plików.

Narzędzia wykrywające przekazy steganograficzne

W sieci (Internet) jest dostępnych wiele narzędzi, które mają służyć do wykrywania przekazów steganograficznych. Przykładowy program do wykrywania przekazów steganograficznych to *Stegdetect* w kilku wersjach, który można pobrać ze strony www.outuess.org. W praktyce są one mało skuteczne, ponieważ aby dany program wykrywający dane był skuteczny, powinien „znać" algorytm programu steganograficznego,

którego działanie ma wykryć, a tzw. programy „uniwersalne" są mało skuteczne. Najpopularniejsze to: *Courier 1.0*, *Image Hide*, *JPHS for Windows 0.5*, *Hide In Picture 2.0*, *Steg Hide 0.5.1*, *Steganography 1.6.5, Stealth Files 4.0 oraz WbStego 4.2*. Pewną metodą wspomagającą wykrywanie przekazów steganograficznych może być zwykłe „przełamanie" haseł użytkownika i kluczy steganograficznych używanych przez znane i powszechnie dostępne programy do ukrywania danych.

Większość metod polega na zbadaniu sposobu zachowania danego programu steganograficznego, prześledzenie jakich zmian dokonuje on w danych nośnych i przeprowadzenie odpowiednich badań statystycznych tych zmian. Okazuje się, że niektóre programy „zostawiają" proste „odciski palców", np:. w postaci zmienionych barw wybranych pikseli w pierwszym rzędzie obrazu. Inne dodają dane na samym końcu pliku tak, że żaden zwykły program ich nie wykryje, jedynie program steganograficzny odbiorcy wiadomości jest w stanie ją odczytać. Inne programy pozostawiają „szum informacyjny". Przykładowy graficzny podgląd zmian, jakie dokonuje przykładowy program steganograficzny w badanym obrazie cyfrowym przedstawia rysunek 5.

Rysunek 5 pokazuje, że graficzna reprezentacja danych bardzo ułatwia wizualną detekcję anomalii. Widać, że barwy obrazu zostały zmienione w sposób losowy. Podobny przykład pokazuje rysunek 7, który uwidacznia zaburzenia w strukturze badanego pliku na jego końcu.

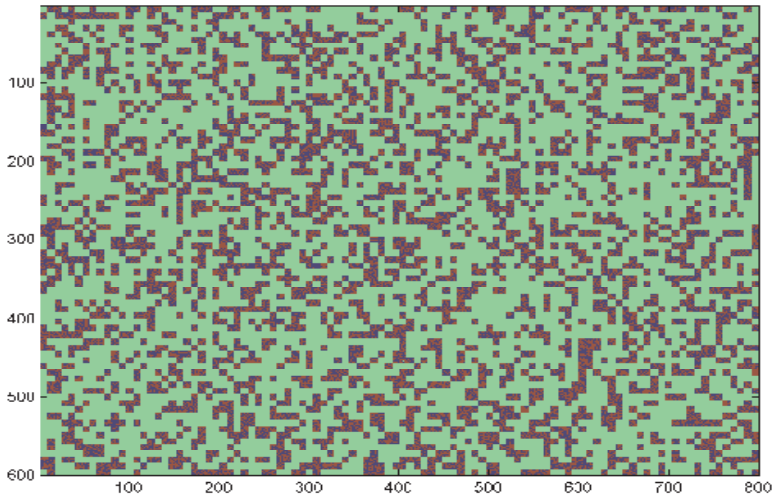
Rysunek 6 przedstawia 5 plików powstałych z oryginału w wyniku ukrycia w nim dodatkowych danych. Widać zmiany, które pojawiają się dopiero od pewnego miejsca w tym pliku. Inny przykład zdemaskowania niepotrzebnych białych znaków w pliku internetowym przedstawia rysunek 7.

Zaznaczenie tekstu w pliku źródłowym strony internetowej może zdemaskować podejrzane ilości bezużytecznych (tylko pozornie) białych znaków (spacja, tabulacja).

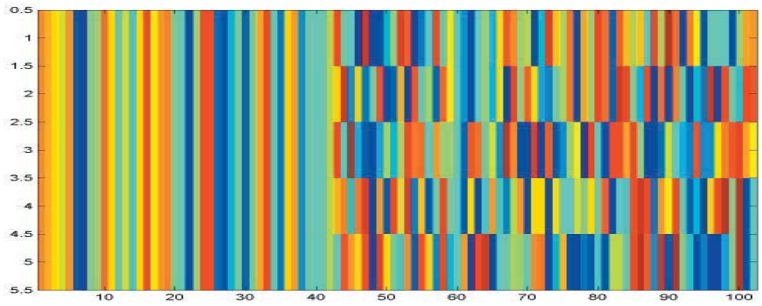
Podsumowując można wysnuć kilka wniosków. W obecnych czasach mamy wielką swobodę w wymianie informacji. Do tego dochodzi ogromny chaos informacyjny i brak konsekwencji w tworzeniu standardów struktur danych. Te czynniki sprzyjają rozwojowi technik ukrywania a także przemycania ogromnych ilości danych.

W sieci Internet jest dostępnych wiele narzędzi służących do ukrywania danych. Za to istniejące narzędzia do wykrywania przekazów steganograficznych, choć jest ich również wiele, praktycznie nie są skuteczne. Do każdego istniejącego programu steganograficznego należałoby stworzyć odpowiednik wykrywający ślady zostawiane przez ten program. Z kolei trudno stworzyć program, który dane te by odczytał. Poza tym, trzeba podkreślić, że do trudności w zdekodowaniu ukrytych danych dochodzi problem, że sam fakt ich ukrycia i lokalizacji jest nieuchwytny.

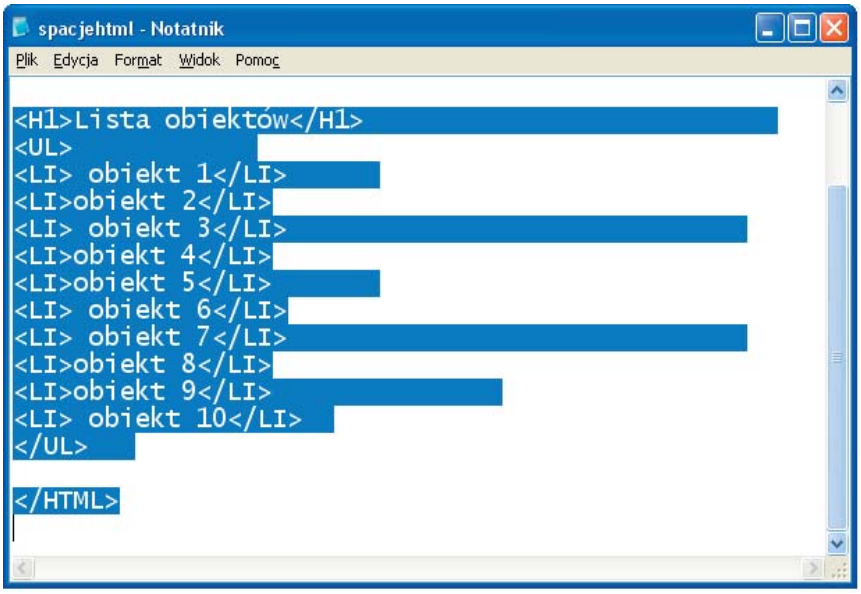
Należy zwrócić większą uwagę na ważność zagrożeń dla bezpieczeństwa informacji, jakie wynikają z korzystania z narzędzi steganograficznych i ukrytych programów. W niektórych krajach programy kryptograficzne są uznawane za pewien rodzaj broni. W naszym kraju oprogramowanie steganograficzne nie jest nigdzie zdefiniowane i ujęte jako część oprogramowania kryptologicznego. Nie trzeba dodawać, że kody źródłowe oprogramowania do steganografii również mogą być ukrywane i przemycane. Można zapobiegać tym zagrożeniom i ustanawiać ściślejsze standardy danych tak, aby nie zostawiać w nich obszarów elastycznych, których użycie nie zostaje zauważone przez autoryzowane programy do ich obsługi.



Rys. 5. Podgląd „różnicy" pomiędzy oryginałem obrazka i jego odpowiednikiem po ukryciu danych



Rys. 6. Podgląd graficzny bajtów na końcu pliku



Rys. 7. Podejrzane białe znaki na końcu linii w źródle pliku witryny internetowej