Steganography

-- hiding in plain sight

David Morgan

What's steganography?

"Covered writing"

steganos = covered*

graphy = writing

*Antonis Christodoulou, an excellent Spring 2008 CS530 student from Greece, challenged this translation. He said there is more to the meaning of the word in Greek than this. But "covered" is how all the English language technical literature presents it.

What's steganography?

- steganography
  - embedding information (plaintext) within other seemingly harmless information (covertext) in such a way that no one but the intended recipient would ***try*** to retrieve it
- *versus* cryptography
  - transforming information (plaintext) into other unintelligible information (ciphertext) such that no one but the intended recipient would ***be able*** to retrieve it

Further differences

- Steganography
  - hide, without altering
  - obfuscates the fact of communication, not the data
  - preventative - deters attacks
- Cryptography
  - alter, without hiding
  - obfuscates the data, not the fact of communication
  - curative - defends attacks

Non-cyber examples

animal camouphlage

Targets are inherent, embedded, camouflaged, implicit in their environment. They blend in with the crowd.

Waldo

Where's Waldo?

Other non-cyber techniques

- subset
- null cipher
- Bacon cipher

Subset

Imagine a package is being prepared for you.

This tells you when and where you can get it:

Dear George;

Greetings to all at Oxford. Many thanks for your letter and for the summer examination package. All Entry Forms and Fees Forms should be ready for final despatch to the Syndicate by Friday 20th or at the very latest, I'm told by the 21st. Admin has improved here, thought there's room for improvement still, just give us all two or three more years and we'll really show you! Please don't let these wretched 16+ proposals destroy your basic O and A pattern. Certainly this sort of change, if implemented immediately, would bring chaos.

Sincerely yours;

Subset

covertext

Dear George;

Greetings to all at Oxford. Many thanks for **your** letter and for the summer examination **package**. All Entry Forms and Fees Forms should be **ready** for final despatch to the Syndicate by **Friday** 20th or at the very latest, I'm told by the **21st**. Admin has improved here, thought there's **room** for improvement still, just give us all two or **three** more years and we'll really show you! **Please** don't let these wretched 16+ proposals **destroy** your basic O and A pattern. Certainly **this** sort of change, if implemented **immediately**, would bring chaos.

Sincerely yours;

11-word message in 93-word covertext
(8.45 ratio – haystack to needle)

plaintext

Null cipher – 1st letter

PRESIDENT'S **E**MBARGO **R**ULING **S**HOULD **H**AVE
IMMEDIATE **N**OTICE. **G**RAVE **S**ITUATION **A**FFECTING
INTERNATIONAL **L**AW. **S**TATEMENT **F**ORESHADOWS
RUIN **O**F **M**ANY **N**EUTRALS. **Y**ELLOW **J**OURNALS
UNIFYING **N**ATIONAL **E**XCITEMENT **I**MMENSELY.

PERSHING SAILS FROM NY JUNE I

24-character message in 204-character covertext (8.50 ratio)

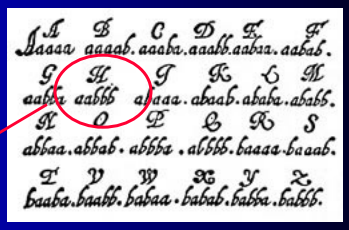
Different covertext, same plaintext

A**P**PARENTLY N**E**UTRAL'S P**R**OTEST I**S** T**H**OROUGHLY
D**I**SCOUNTED A**N**D I**G**NORED. I**S**MAN H**A**RD H**I**T.
B**L**OCKADE I**S**SUE A**F**FECTS P**R**ETEXT F**O**R E**M**BARGO
O**N** B**Y**PRODUCTS, E**J**ECTING S**U**ETS A**N**D V**E**GETABLE
O**I**LS.

PERSHING SAILS FROM NY JUNE I

24-character message in 176-character covertext (7.33 ratio)

Bacon's cipher



H a v e f u n

aabbb aaaaa baabb aabaa aabab baabb abbaa

BUrgeR WITH fRIes TAsTY BUt Not FOr hEalTH

uses a "bilateral" alphabet
each letter has 2 possible fonts (or cases)

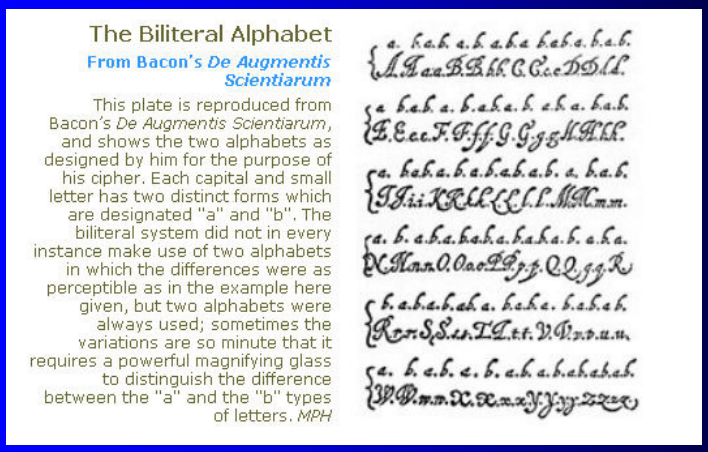
7-character message in 35-character covertext (5.00 ratio)

What's this one?

USc atHlETICS is SURpasSed BY ComPuTer ScIenCE

Hint: starts with same letter as previous because BUrge == USCat

A less obvious bilateral alphabet

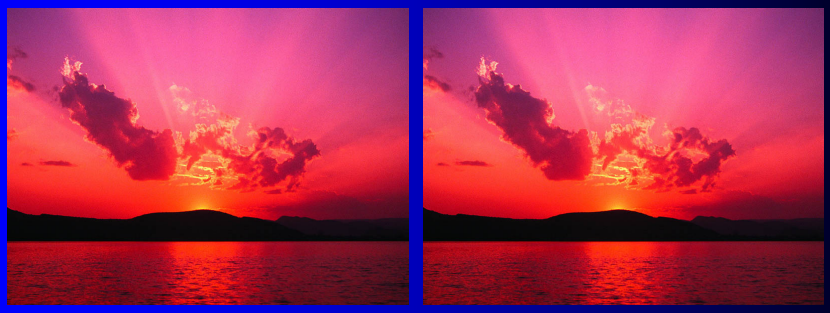


from The Philosophical Research Society at http://www.prs.org/gallery-bacon.htm

Doing it with computers

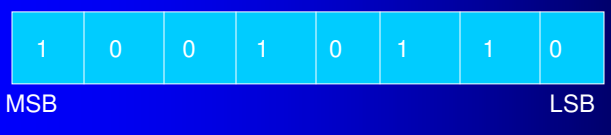
- Steganography – hiding a file inside of another
  - typically hiding text inside of a media file
  - normally used for the transportation of secretive information
- Operating System
  - unused memory
    - slack space
    - unallocated space
  - hidden partition
  - normally used to hide data from investigators
- Network
  - unused bits in packet headers
  - spread spectrum, frequency shifting

Photo as cover - any difference?



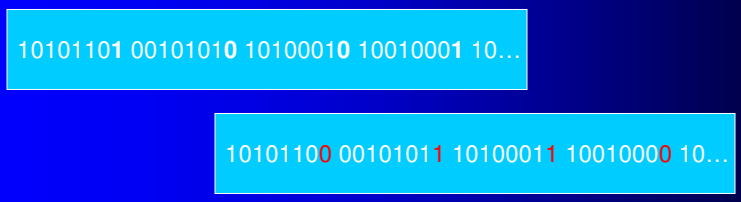
Least Signifiacant Bit Manupalation

- Idea is that the least significant bit of a byte can change with little change to the overall file
- Consider a 8-bit grey scale image
  - One pixel of information is stored using 8 bits.
  - There are 256 different variations of grey.



LSB continued

- Change in the LSB information of some area of the image will not be noticeable by naked eye.
- Utilizing this fact the message is embedded



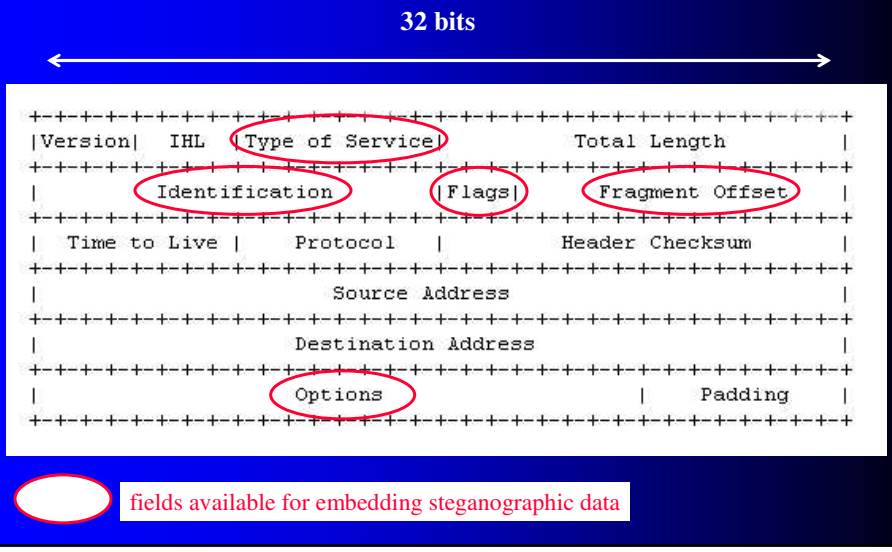
LSB advantages and disadvantages

- Advantages
  - Does not change the size of the file
  - Is harder to detect than other steganography techniques
- Disadvantages
  - Normally must use the original program to hide and reveal data
  - If the picture with the hidden information is converted to another format, then the hidden data may be lost

Some network examples

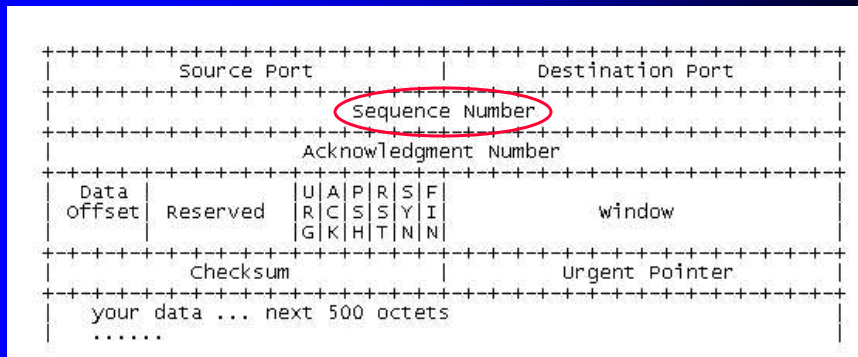
- ***embedding*** data directly
  - in header fields, and/or
  - in payload
- ***expressing*** data by network event timing
  - data is just patterns
  - can be non-material
  - e.g., morse code

IP packet header



TCP packet (segment) header

32 bits



Put 'em where they don't belong

because you can

*fields available for embedding steganographic passengers

The protocols don't restrict

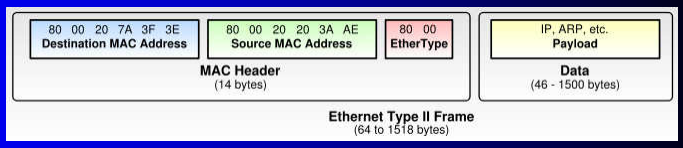
- IP "identification" field's value
  - "An internet header field carrying the identifying value assigned by the sender to aid in assembling the fragments of a datagram."
    RFC 791, "Internet Protocol"
- TCP "sequence number" field's value
  - "When new connections are created, an initial sequence number (ISN) generator is employed which selects a new 32 bit ISN. The generator is bound to a ... clock ... [but] not tied to a global clock in the network, and TCPs may have different mechanisms for picking the ISN's."
    RFC 793, Transmission Control Protocol

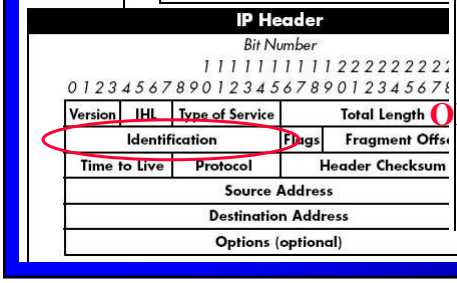
Proof-of-concept covert channel demo

- Named "covert_tcp" by Craig Rowland
- client/sender and server/receiver roles
- client places data in either
  - IP header's "identification" field, or
  - TCP header's "sequence number" field
- server knows, fetches the data out

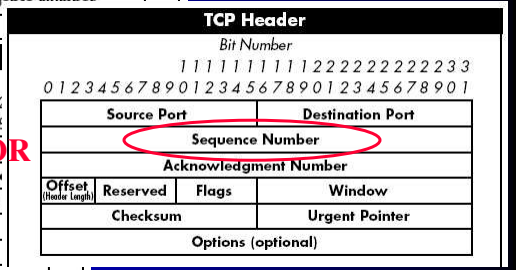
http://www.firstmonday.org/Issues/issue2_5/rowland/

Fields alternatively utilized





OR



Simultaneous screenshots

client/sender (on 192.168.1.20)

```
[root@V1 root]# ./covert_tcp  -dest 192.168.1.132 -source 192.168.1.20 -source_port 1234 -dest_port 80 -file covert_data_to_send
Covert TCP 1.0 (c)1996 Craig H. Rowland
(crowland@psionic.com)
Not for commercial use without permission.
Destination Host: 192.168.1.132
Source Host     : 192.168.1.20
Originating Port: 1234
Destination Port: 80
Encoded Filename: covert_data_to_send
Encoding Type   : IP ID

Client Mode: Sending data.

Sending Data: A
Sending Data: B
Sending Data: C
[root@V1 root]
```

file content: ABC

server/receiver (on 192.168.1.132)

```
[root@clay ~]# ./covert_tcp -server -dest 192.168.1.132 -source 192.168.1.20 -file captured_data.txt
Covert TCP 1.0 (c)1996 Craig H. Rowland
(crowland@psionic.com)
Not for commercial use without permission.
Listening for data from IP: 192.168.1.20
Listening for data bound for local port: Any Port
Decoded Filename: captured_data.txt
Decoding Type Is: IP packet ID

Server Mode: Listening for data.

Receiving Data: A
Receiving Data: B
Receiving Data: C
```

Packet dump seen at server

-- using IP identification field

```
[root@localhost ~]# tcpdump -nntv -r ipid.server.cap  | cat -n
reading from file ipid.server.cap, link-type EN10MB (Ethernet)
     1  IP (tos 0x0, ttl  64, id 16640, offset 0, flags [none], proto: TCP (6), length: 40) 192.168.1.20.1234 > 192.168.1.132
     2  IP (tos 0x0, ttl  64, id 0, offset 0, flags [DF], proto: TCP (6), length: 40) 192.168.1.132.80 > 192.168.1.20.1234: R
     3  IP (tos 0x0, ttl  64, id 16896, offset 0, flags [none], proto: TCP (6), length: 40) 192.168.1.20.1234 > 192.168.1.132
     4  IP (tos 0x0, ttl  64, id 0, offset 0, flags [DF], proto: TCP (6), length: 40) 192.168.1.132.80 > 192.168.1.20.1234: R
     5  IP (tos 0x0, ttl  64, id 17152, offset 0, flags [none], proto: TCP (6), length: 40) 192.168.1.20.1234 > 192.168.1.132
     6  IP (tos 0x0, ttl  64, id 0, offset 0, flags [DF], proto: TCP (6), length: 40) 192.168.1.132.80 > 192.168.1.20.1234: R
[root@localhost ~]#
```

Letter	Ascii code	
A	65	65 x 256 = 16640
B	66	66 x 256 = 16896
C	67	67 x 256 = 17152
D	68	
etc	etc	

Simultaneous screenshots

client/sender (on 192.168.1.20)

```
[root@V1 root]# ./covert_tcp -seq -dest 192.168.1.132 -source 192.168.1.20 -source_port 1234 -dest_port 80 -file covert_data_to_send
Covert TCP 1.0 (c)1996 Craig H. Rowland
(crowland@psionic.com)
Not for commercial use without permission.
Destination Host: 192.168.1.132
Source Host     : 192.168.1.20
Originating Port: 1234
Destination Port: 80
Encoded Filename: covert_data_to_send
Encoding Type   : IP Sequence Number

Client Mode: Sending data.

Sending Data: A
Sending Data: B
Sending Data: C
[root@V1 root]
```

server/receiver (on 192.168.1.132)

```
[root@clay ~]# ./covert_tcp -seq -server -dest 192.168.1.132 -source 192.168.1.20 -file captured_data.txt
Covert TCP 1.0 (c)1996 Craig H. Rowland
(crowland@psionic.com)
Not for commercial use without permission.
Listening for data from IP: 192.168.1.20
Listening for data bound for local port: Any Port
Decoded Filename: captured_data.txt
Decoding Type Is: IP Sequence Number

Server Mode: Listening for data.

Receiving Data: A
Receiving Data: B
Receiving Data: C
```

Packet dump seen at server

-- using TCP sequence number field

```
[root@localhost ~]# tcpdump -nntA -r seq.server.cap  | cat -n
reading from file seq.server.cap, link-type EN10MB (Ethernet)
     1  IP 192.168.1.20.1234 > 192.168.1.132.80: S 1090519040:1090519040(0) win 512
     2  E..(....@.............PA..... ....P.............
     3  IP 192.168.1.132.80 > 192.168.1.20.1234: R 0:0(0) ack 1090519041 win 0
     4  E..(..@.@............P......A...P.......
     5  IP 192.168.1.20.1234 > 192.168.1.132.80: S 1107296256:1107296256(0) win 512
     6  E..(G...@.............PB..... ....P.............
     7  IP 192.168.1.132.80 > 192.168.1.20.1234: R 0:0(0) ack 16777217 win 0
     8  E..(..@.@............P......B...P.......
     9  IP 192.168.1.20.1234 > 192.168.1.132.80: S 1124073472:1124073472(0) win 512
    10  E..(....@.............PC..... ....P.............
    11  IP 192.168.1.132.80 > 192.168.1.20.1234: R 0:0(0) ack 33554433 win 0
    12  E..(..@.@............P......C...P.......
[root@localhost ~]#
```

Voice stream as cover

- invent your own protocol

Table 1. Header fields and their function

Type of field	No. of bits	Function
P (Parameter)	4	Indicates parameter that is transmitted inside the watermark
S (Side)	1	Indicates the side of the communication (1 - sender, 0 – receiver)
C (Continuity)	1	Describes if a packet contains the beginning or continuation of the parameter indicated in the field P (1 – beginning of new parameter, 0 – continuation of the last parameter)

Exemplary values of the field P are shown below:
0001 – authentication or integrity parameter (32 bits)
0010 – informational parameter 1 (32 bits)
0011 – informational parameter 2 (32 bits)
0100 – informational parameter 3 (32 bits)
0101 – post authentication and integrity parameter (32 bits)
…

"Covert channel for improving VoIP security"
http://www.ippt.gov.pl/~zkotulsk/Covert%20Channel%20for%20Improving%20VoIP%20Security.pdf

Voice stream as cover

- embed your protocol in VoIP packets

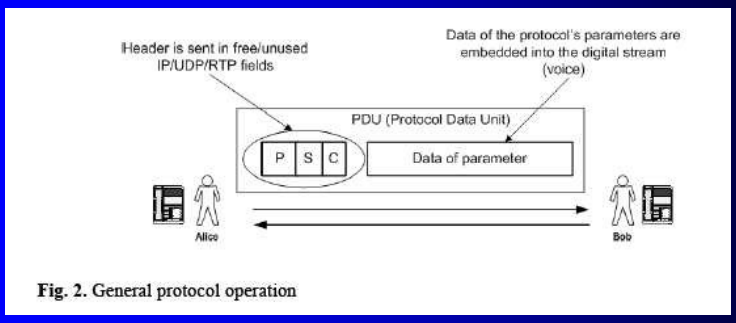


Fig. 2. General protocol operation

"Covert channel for improving VoIP security"
http://www.ippt.gov.pl/~zkotulsk/Covert%20Channel%20for%20Improving%20VoIP%20Security.pdf

Photo, voice work well because … bad is good enough

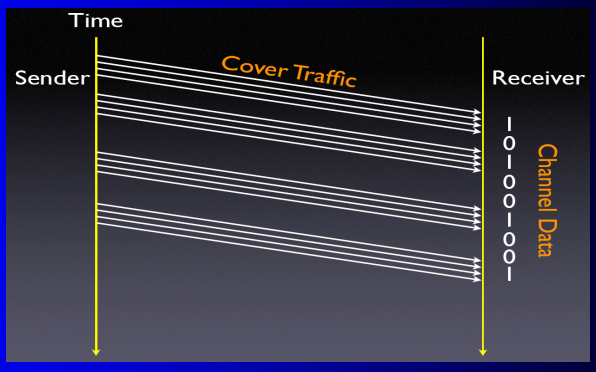
- for human consumption
- our crude senses have high "error" tolerance
- right and wrong – mom never taught me the difference
  - (slightly) wrong colors look the same as right
  - (slightly) wrong voice sounds the same as right

Even more covert…

- signal timing
- port knocking

Even more covert…

- Embed nothing in the non-surreptitious channel
- Instead express covert info by timing channel's non-covert exchanges



http://userpages.umbc.edu/~chauhan2/CMSC691I/Embedding_Covert_channels_into_TCP_IP.ppt

Port knocking…

- encode information in sequence of “open port” requests
  - the port sequence *is* the information e.g.
    - port 72, port 69, port 76, port 76, port 79 requests, signify “hello”
  - the port sequence is a password/combination to “unlock” a response e.g.
    - port 10004, port 10030, port 10012, signify “turn on httpd on port 80”
- receiver observes, optionally reacts
  - daemon sniffs network
  - process watches firewall log

http://www.portknocking.org/

Products

- s-tools
- outguess
- various others
  - http://www.jjtc.com/Security/stegtools.htm
  - http://caia.swin.edu.au/cv/szander/cc/cc-implementations-bib.html