A Picture is Worth a Thousand Words, Literally

Deep Neural Networks for Social Stego

Philip Tully | Mike Raggo

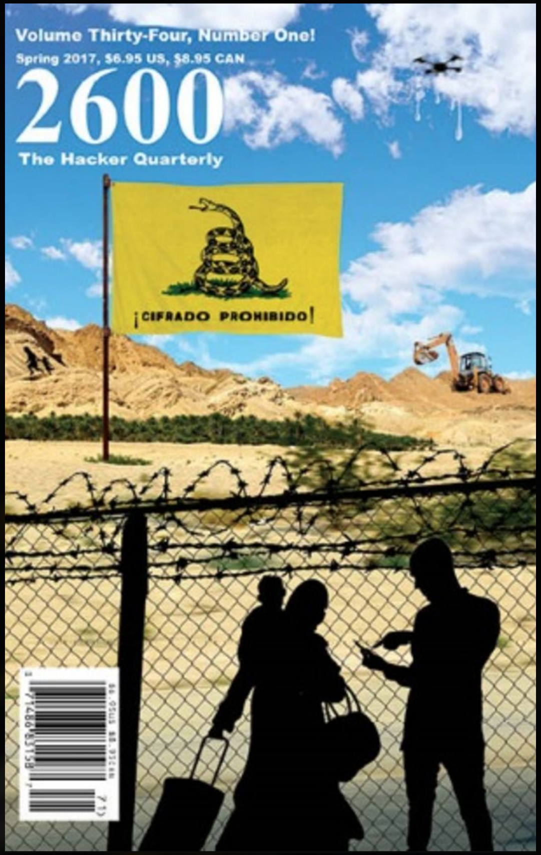
DC25: Community, Discovery and the Unintended Uses of Technology

2600: The Hacker Quarterly

2600
Secret Service
CONFIDENTIAL
The Hacker Quarterly
VOLUME NINE, NUMBER TWO
SUMMER 1992

Summer 1992

25 years


Volume Thirty-Four, Number One!
Spring 2017, $6.95 US, $8.95 CAN
2600
The Hacker Quarterly
¡CIFRADO PROHIBIDO!

Summer 2017

#whoami

Philip Tully

@phtully

ZEROFOX®

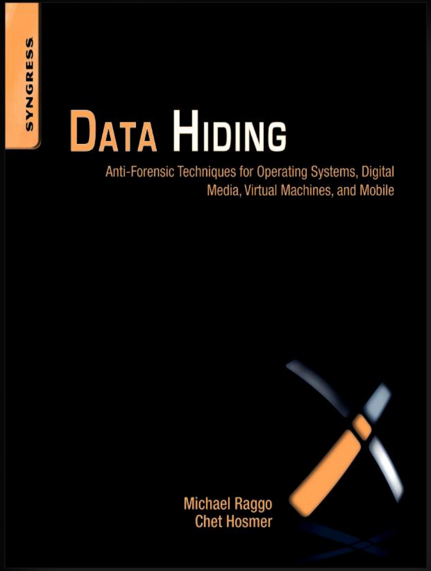
Principal Data Scientist at ZeroFOX

PhD (KTH & University of Edinburgh)

Machine Learning and Neural Nets

Mike Raggo

@datahiding



SYNGRESS

DATA HIDING

Anti-Forensic Techniques for Operating Systems, Digital Media, Virtual Machines, and Mobile

Michael Raggo
Chet Hosmer

802secure

CSO @802 Secure, 17 yrs Stego Research

StegSpy DC12, Author “Data Hiding”

NSA National Cryptologic Museum

The Evolution of Steganography ●

DIY Social Steganography ●

Deep Neural Networks for Social Stego ●

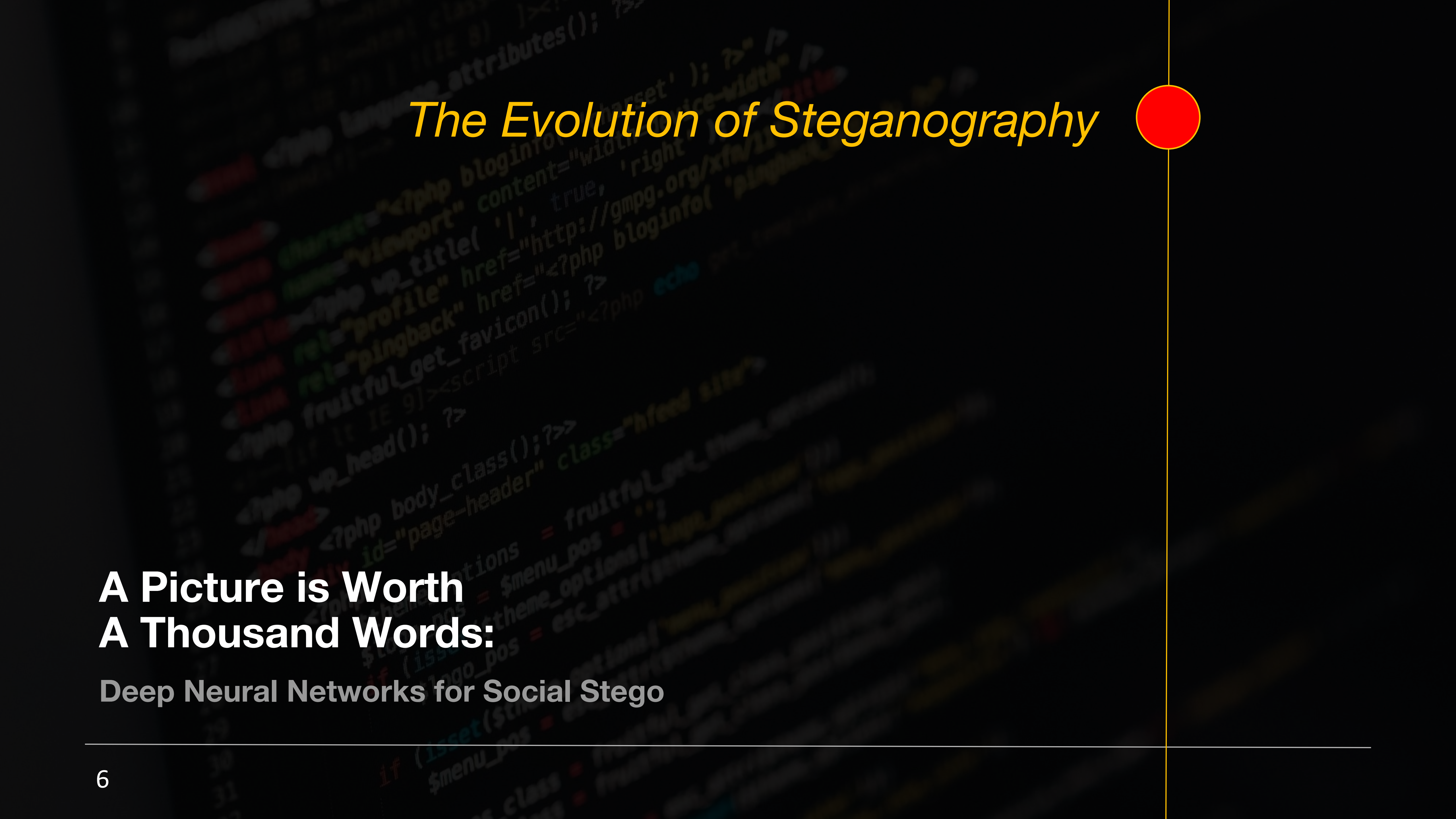
Data-Driven Red and Blue Teaming ●

Wrap Up ●

A Picture is Worth A Thousand Words:

Deep Neural Networks for Social Stego

The Evolution of Steganography



A Picture is Worth A Thousand Words:

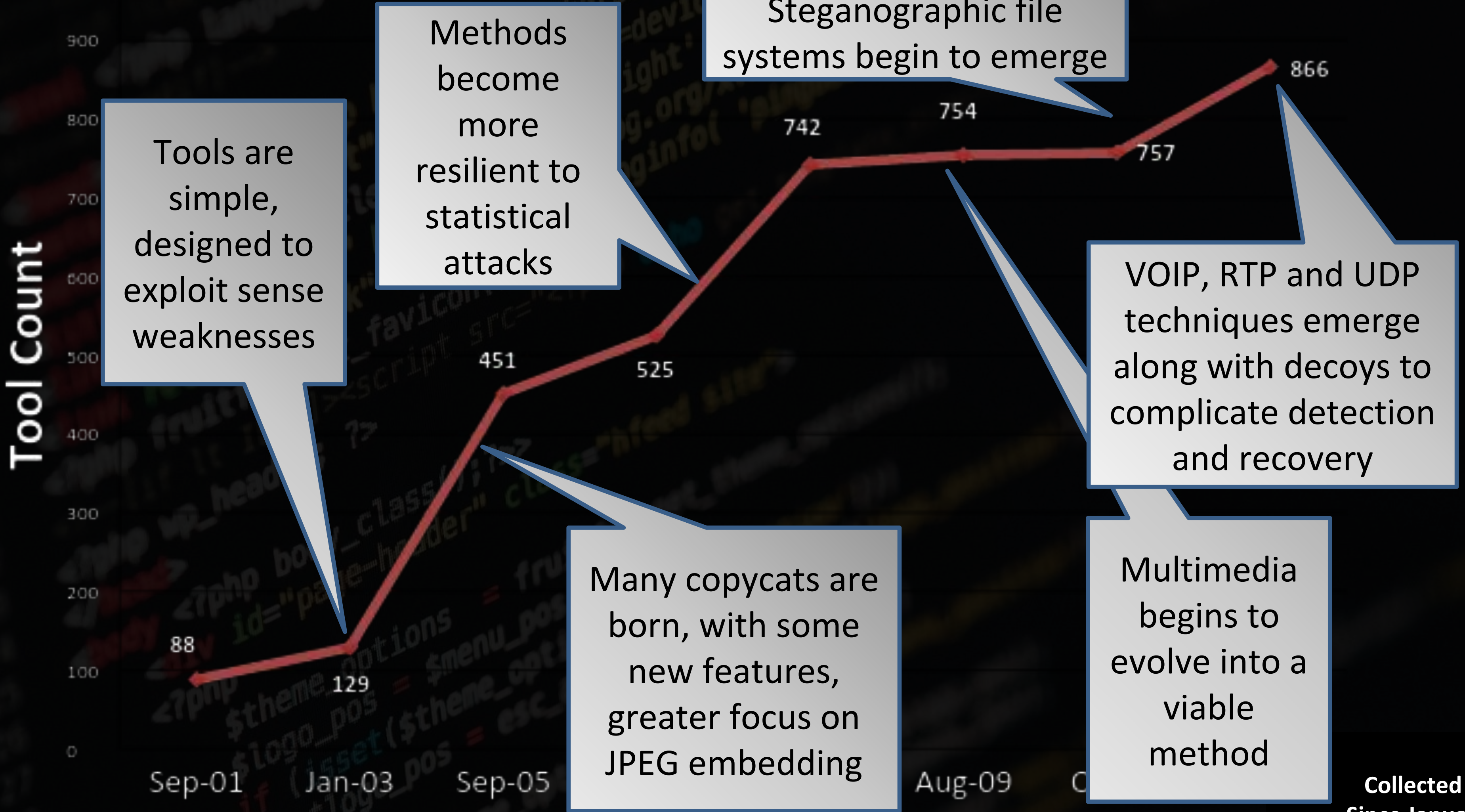
Deep Neural Networks for Social Stego

Covert Communication

“. . . any communication channel that can be exploited by a process to transfer information in a manner that violates the system's security policy.”

Source: U.S. Department of Defense. Trusted Computer System Evaluation “The Orange Book”. Publication DoD 5200.28-STD. Washington: GPO 1985

Evolution of Methods



Tools are simple, designed to exploit sense weaknesses

Methods become more resilient to statistical attacks

Many copycats are born, with some new features, greater focus on JPEG embedding

New versions emerge of existing Alg. Steganographic file systems begin to emerge

Multimedia begins to evolve into a viable method

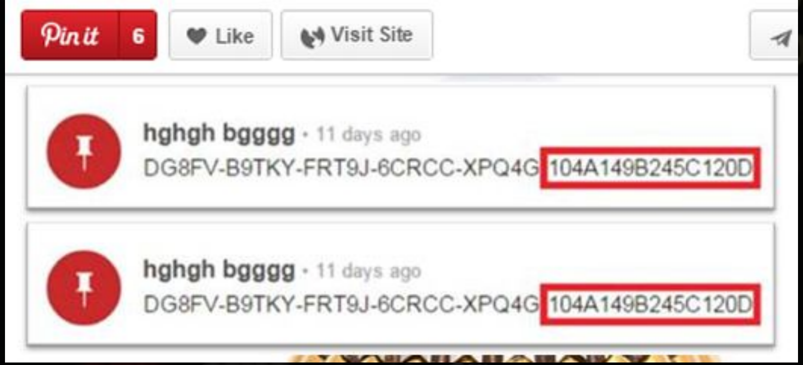
VOIP, RTP and UDP techniques emerge along with decoys to complicate detection and recovery

WetStone Labs
Collected Steganography Programs
Since January 1999 Includes versions

Evolution of Stego in the Internet Era

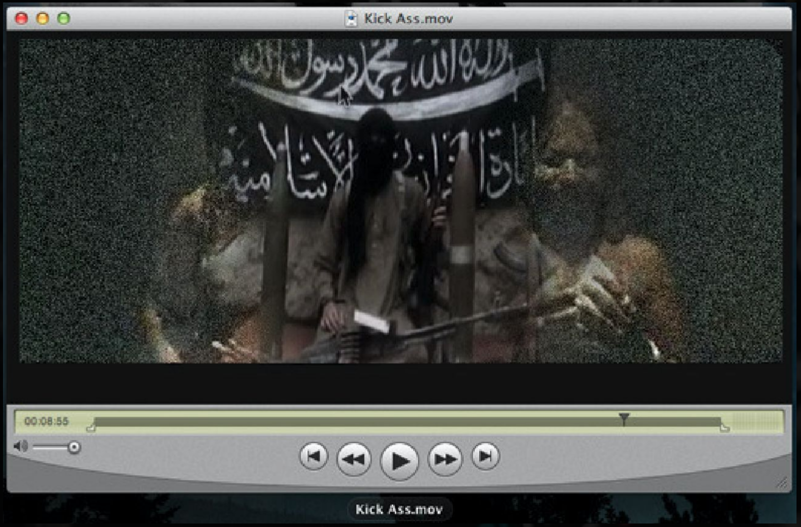
- Stego Apps Decoy Techniques (OpenPuff)
- Stealth Alternate Data Streams (NT)
- Weaponized CnC - Operation Shady RAT (McAfee)
- Prootocols - VOIP, RTP, UDP => WiFi StegoStuffing, Bluetooth (Hosmer/Raggo - Wall of Sheep/Skytalks DC23 & 24)
- MP3 ID3 Metadata exploitation - Hosmer/Raggo Skytalks DC24
- SmartWatch SWATtackhide.py Tizen SDK - Mike Raggo - DEF CON 24 Demo Labs and Wall of Sheep

Types of Steganography

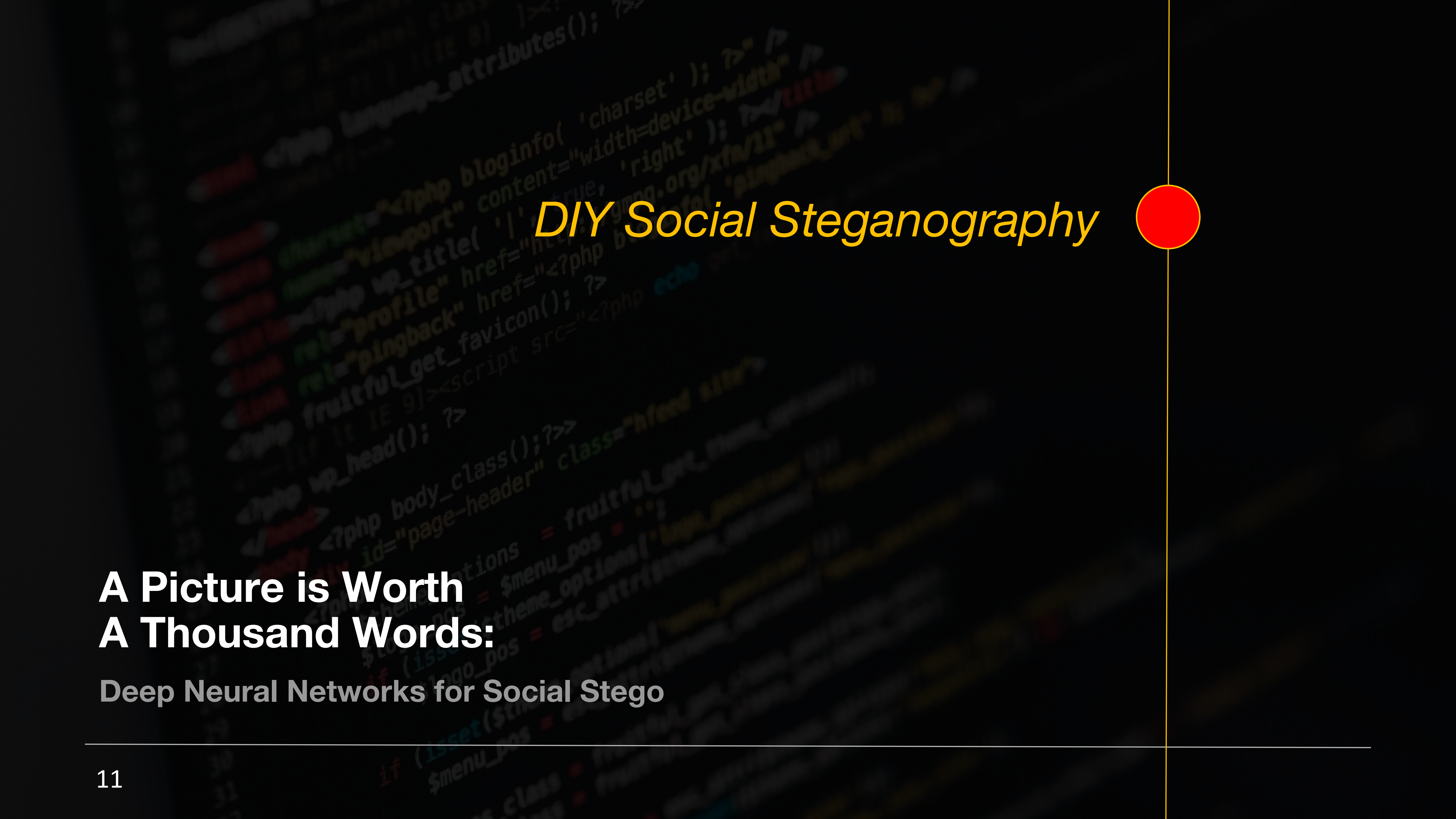


hghgh bgggg · 11 days ago
DG8FV-B9TKY-FRT9J-6CRCC-XPQ4G104A149B245C120D

hghgh bgggg · 11 days ago
DG8FV-B9TKY-FRT9J-6CRCC-XPQ4G104A149B245C120D



- Text/Linguistic Stego - using Natural Language
- Image
  - Spatial (e.g. LSB)
  - Frequency (DCT/DWT)
  - Metadata (varies by file type and versions) - JPEG EXIF vs. JFIF
- Audio
- Video
- Protocols
- Use of crypto with stego
  - Vigenere, base64, XOR, etc.



DIY Social Steganography

A Picture is Worth A Thousand Words:

Deep Neural Networks for Social Stego

Signals in the Social Noise

4.75 billion
pieces of content
shared per day.

80+ million
images uploaded per
day.

100+ hours
of video uploaded
per minute.

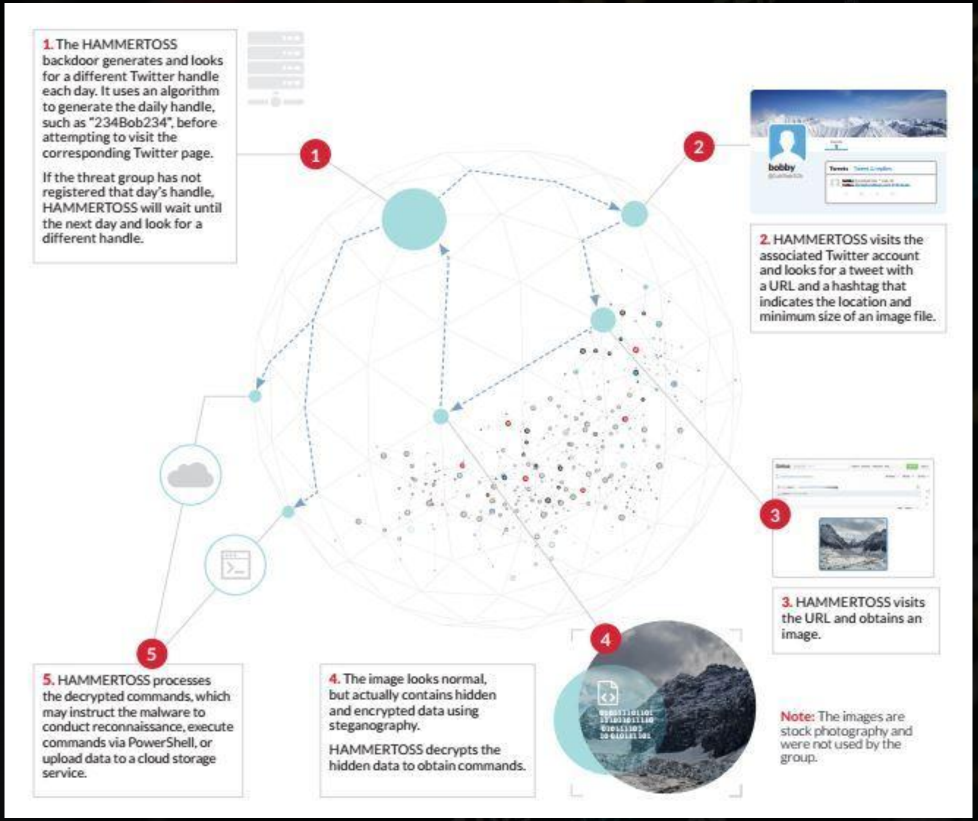
5 billion
+1’s per day.

500+ million
tweets per day.

Social Network Image Proliferation

- Image-based social networks have the fastest growing user bases
- Image-based social networks enjoy the highest daily time spent by users
- “Photos or Images” is the content category most frequently share by users
- Social posts containing images produce 650% higher engagement than text alone

Social Stego in the Wild



1. The HAMMERTOSS backdoor generates and looks for a different Twitter handle each day. It uses an algorithm to generate the daily handle, such as "234Bob234", before attempting to visit the corresponding Twitter page.

If the threat group has not registered that day's handle, HAMMERTOSS will wait until the next day and look for a different handle.

2. HAMMERTOSS visits the associated Twitter account and looks for a tweet with a URL and a hashtag that indicates the location and minimum size of an image file.

3. HAMMERTOSS visits the URL and obtains an image.

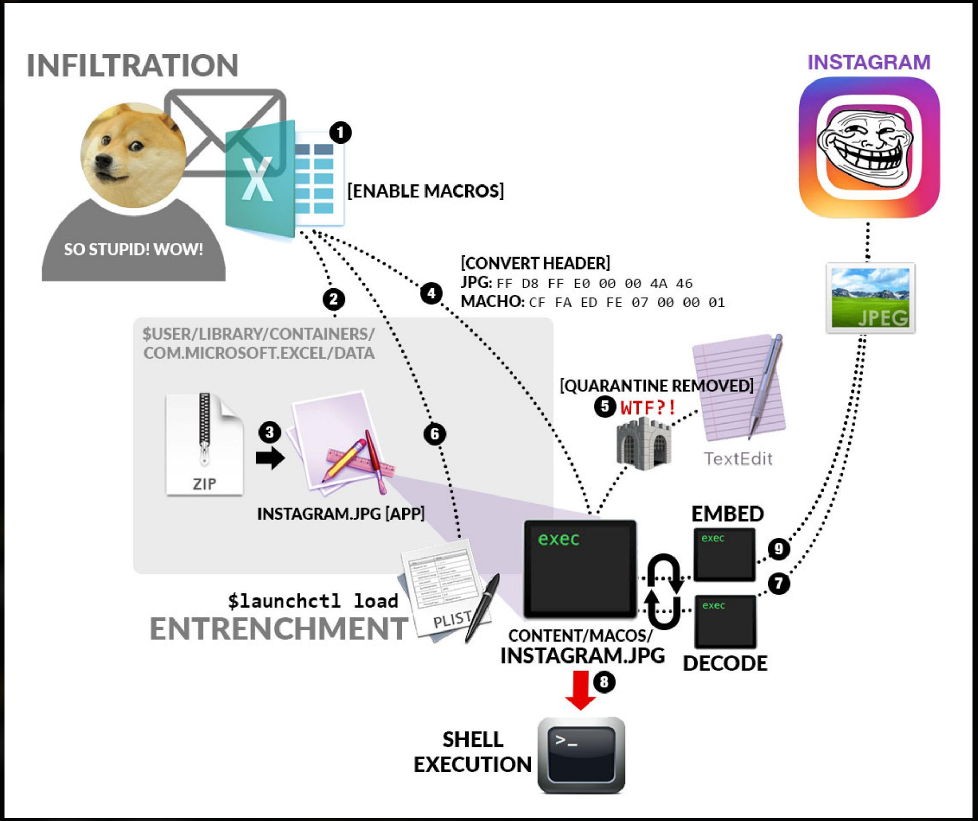
4. The image looks normal, but actually contains hidden and encrypted data using steganography.

HAMMERTOSS decrypts the hidden data to obtain commands.

5. HAMMERTOSS processes the decrypted commands, which may instruct the malware to conduct reconnaissance, execute commands via PowerShell, or upload data to a cloud storage service.

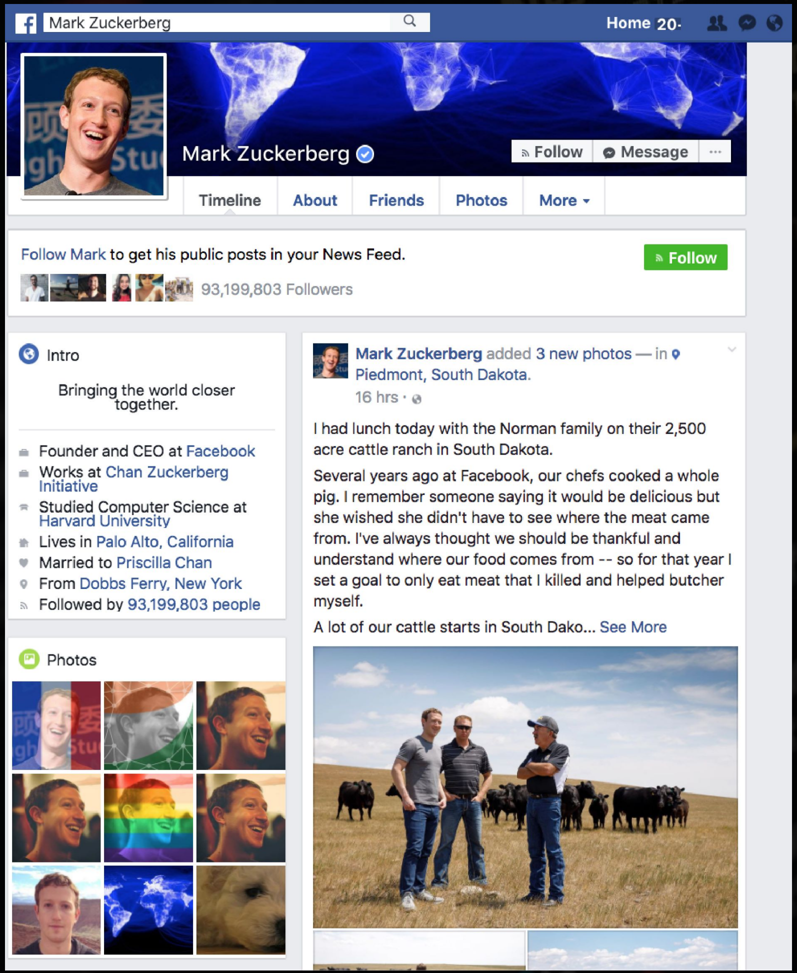
Note: The images are stock photography and were not used by the group.

Black Hat: HAMMERTOSS



White Hat: Instegogram

Social Network Photo Targets



Mark Zuckerberg

Follow Mark to get his public posts in your News Feed.

93,199,803 Followers

Bringing the world closer together.

- Founder and CEO at Facebook
- Works at Chan Zuckerberg Initiative
- Studied Computer Science at Harvard University
- Lives in Palo Alto, California
- Married to Priscilla Chan
- From Dobbs Ferry, New York
- Followed by 93,199,803 people

Mark Zuckerberg added 3 new photos — in Piedmont, South Dakota.
16 hrs ·

I had lunch today with the Norman family on their 2,500 acre cattle ranch in South Dakota.

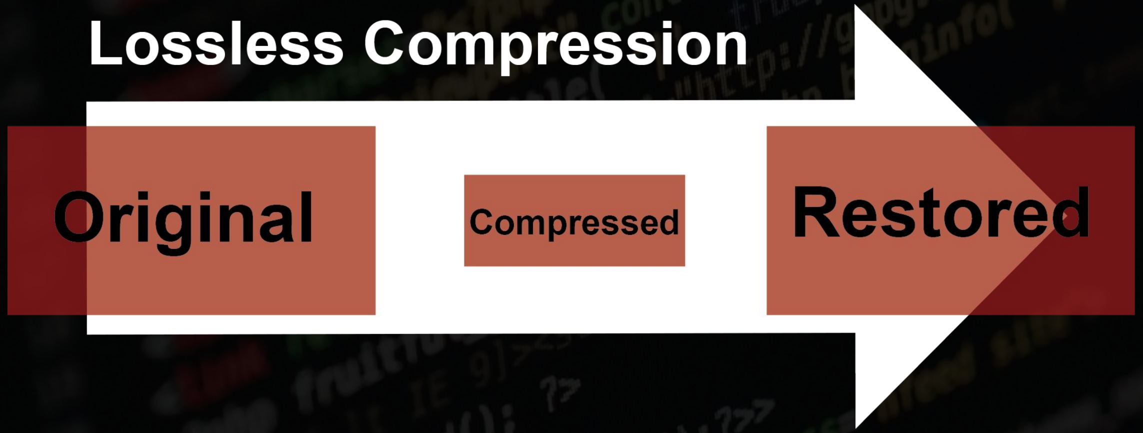
Several years ago at Facebook, our chefs cooked a whole pig. I remember someone saying it would be delicious but she wished she didn't have to see where the meat came from. I've always thought we should be thankful and understand where our food comes from -- so for that year I set a goal to only eat meat that I killed and helped butcher myself.

A lot of our cattle starts in South Dako... See More

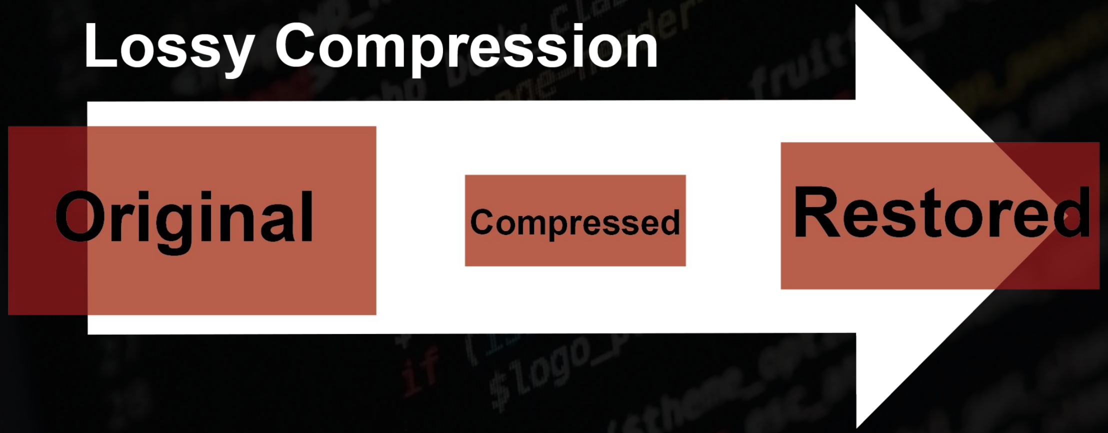
- Profile Image
- Background Image
- Posted Image(s)
- Photo albums
- DM images

Carrier Image File Types

Lossless Compression



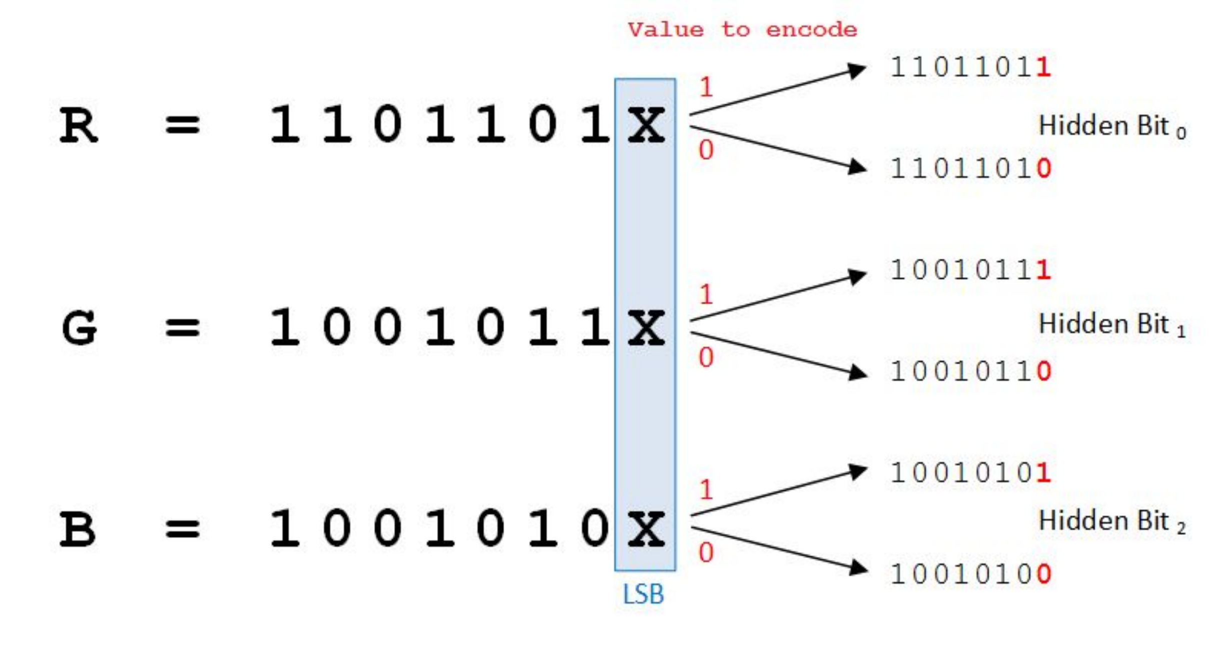
Lossy Compression



- Image quality properties:
  - Lossy v. Lossless Raster Compression
  - DPI/PPI

- Common file formats:
  - JPEG (Lossy)
  - PNG (Lossless)
  - TIFF (Lossless)
  - GIF (Lossy)
  - BMP (Lossy)

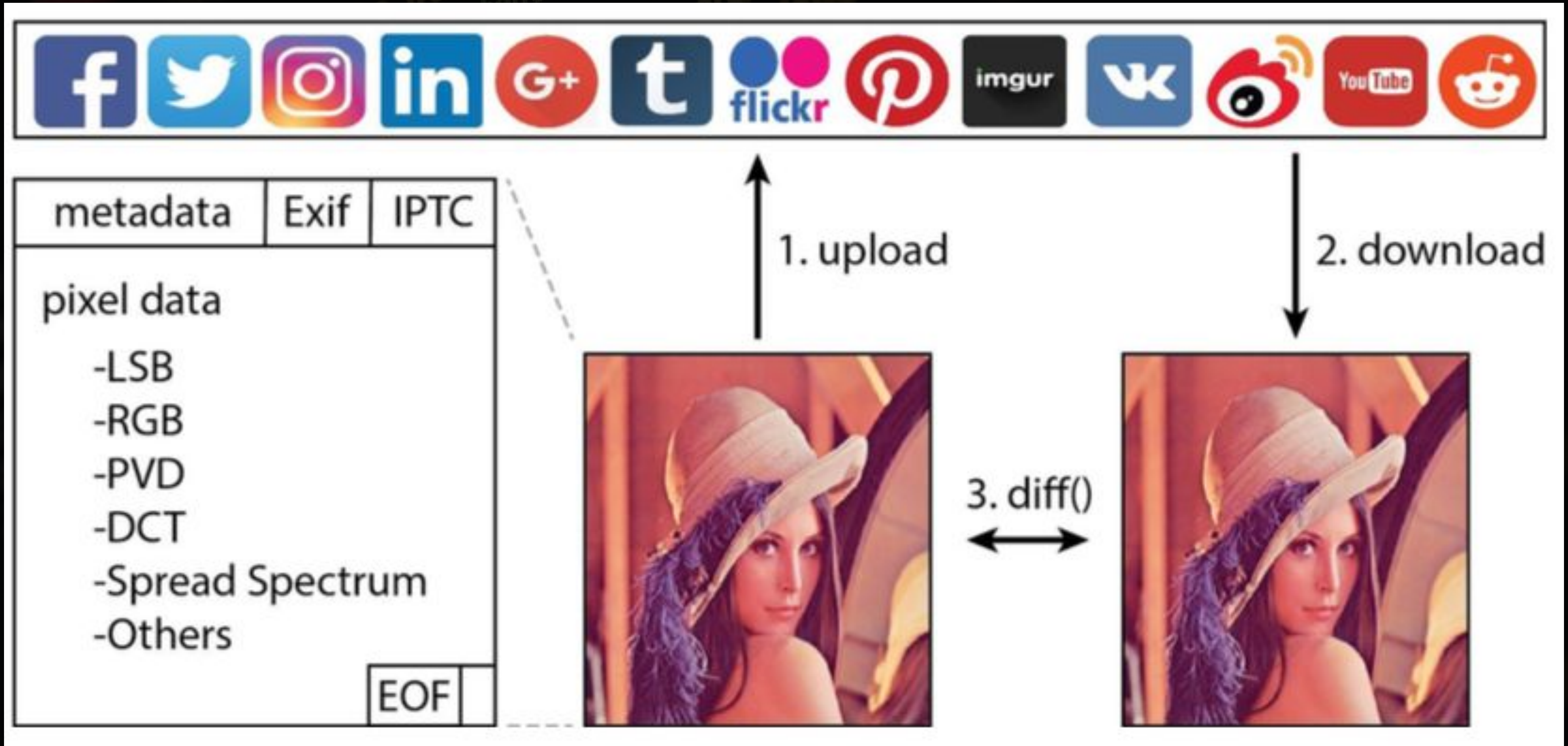
Trial and Error - Attempted Methods



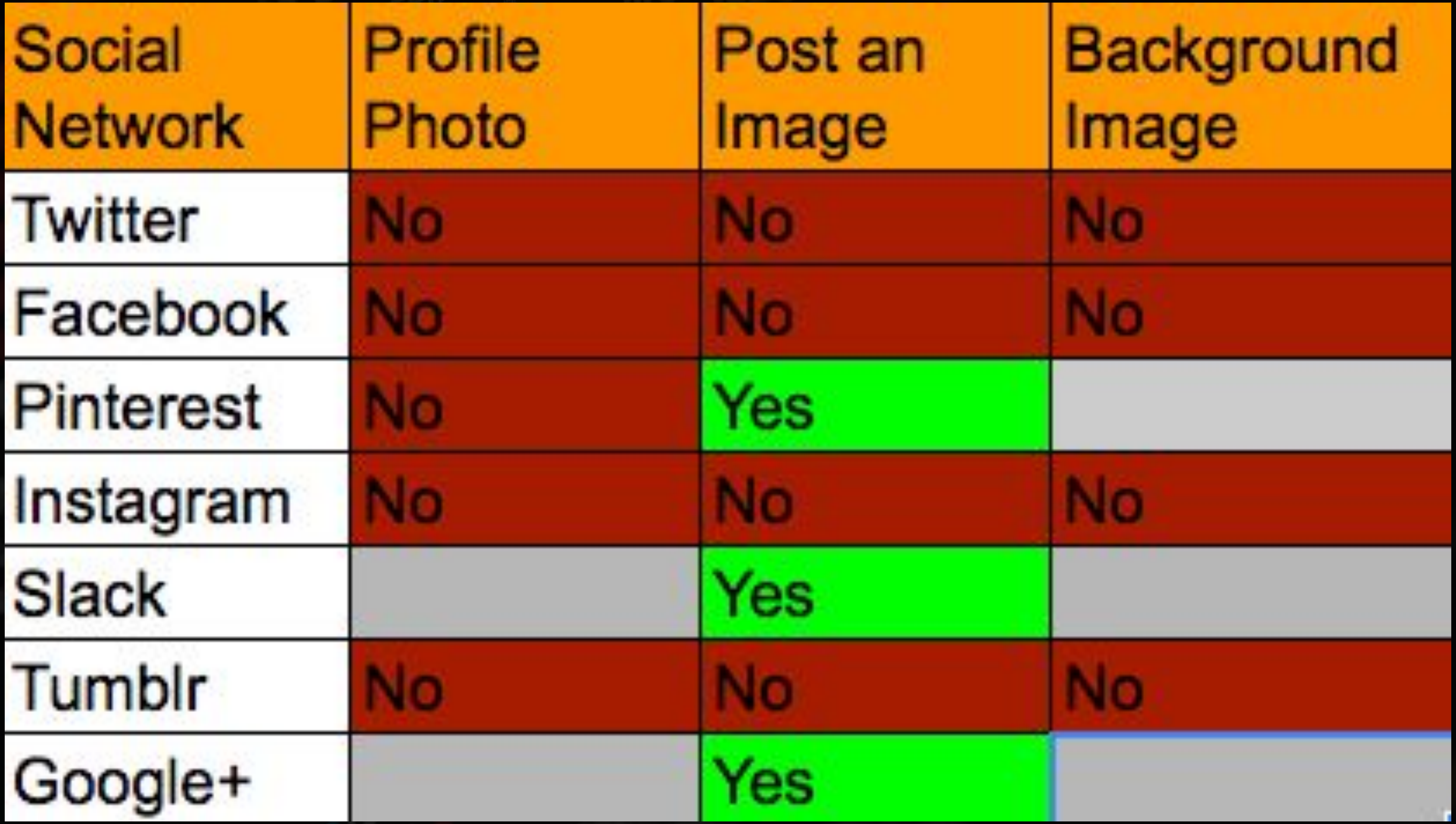
DataGenetics

- Metadata fields (varies by image types JPEG EXIF vs. JFIF, etc.)
- LSB - Least Significant Bit
- Insertion
- Append after EOF marker
- Pre and Post Upload
- Linguistic Steganography

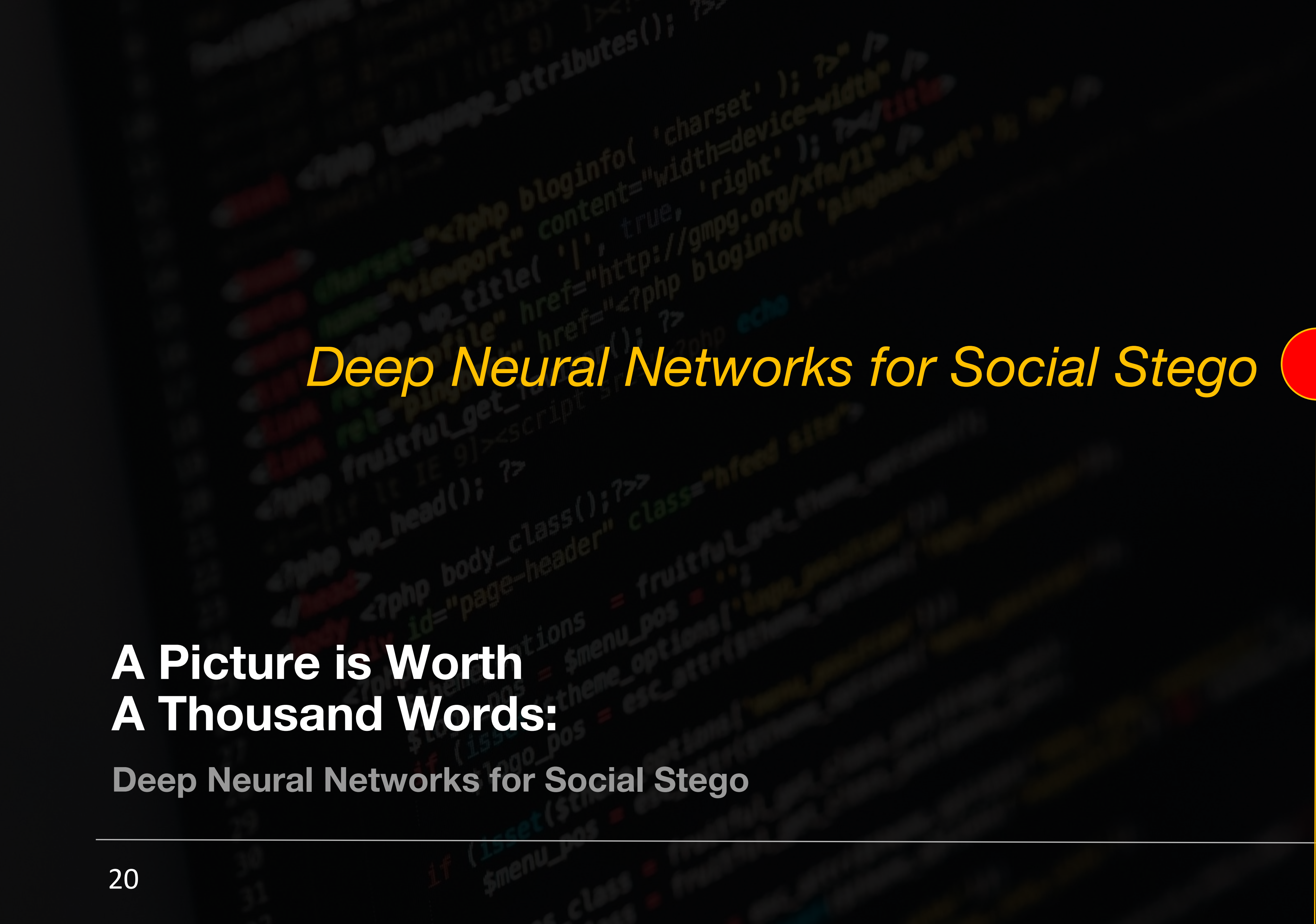
High-Level Testing Workflow



Social Network Data Hiding Survivability



Social Network	Profile Photo	Post an Image	Background Image
Twitter	No	No	No
Facebook	No	No	No
Pinterest	No	Yes	
Instagram	No	No	No
Slack		Yes	
Tumblr	No	No	No
Google+		Yes	

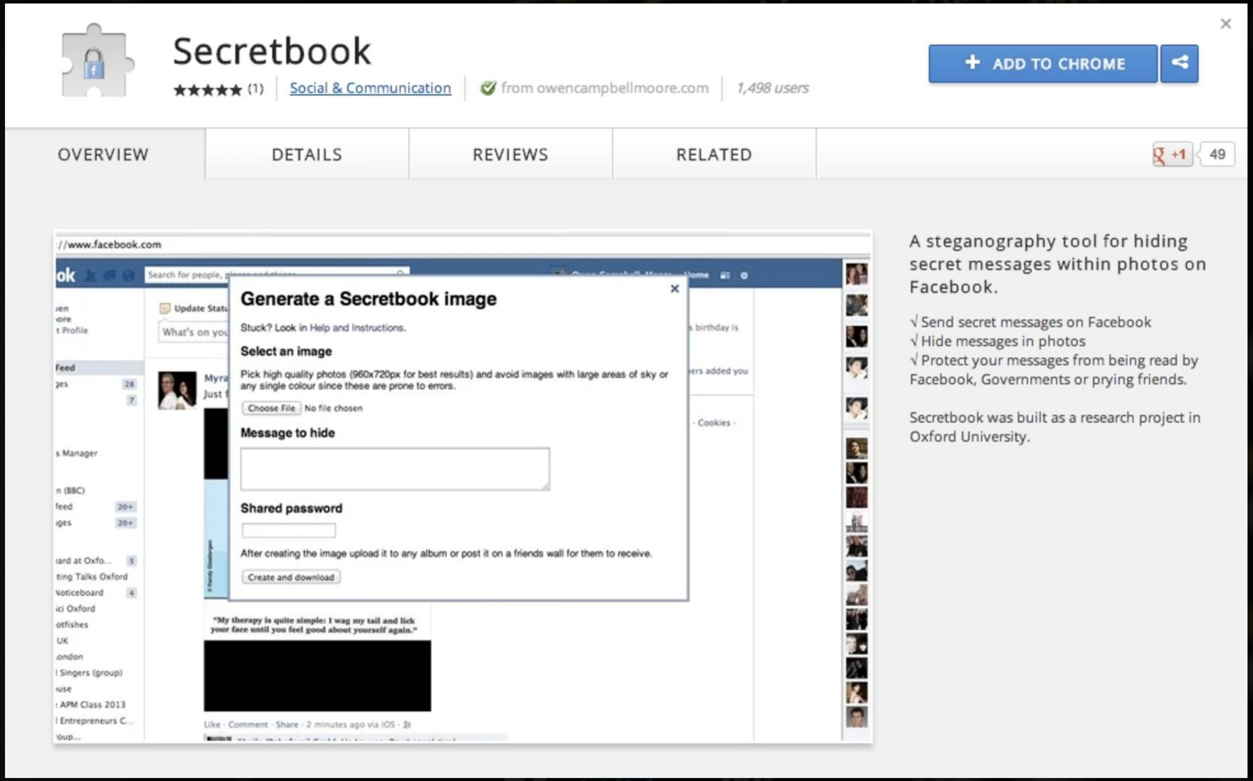


Deep Neural Networks for Social Stego

A Picture is Worth
A Thousand Words:

Deep Neural Networks for Social Stego

Secretbook by Owen Campbell-Moore



Secretbook

★★★★★ (1) | Social & Communication | from owencampbellmoore.com | 1,498 users

Generate a Secretbook image

Stuck? Look in Help and Instructions.

Select an image

Pick high quality photos (960x720px for best results) and avoid images with large areas of sky or any single colour since these are prone to errors.

Choose File No file chosen

Message to hide

Shared password

After creating the image upload it to any album or post it on a friends wall for them to receive.

Create and download

"My therapy is quite simple: I wag my tail and lick your face until you feel good about yourself again."

[redacted]

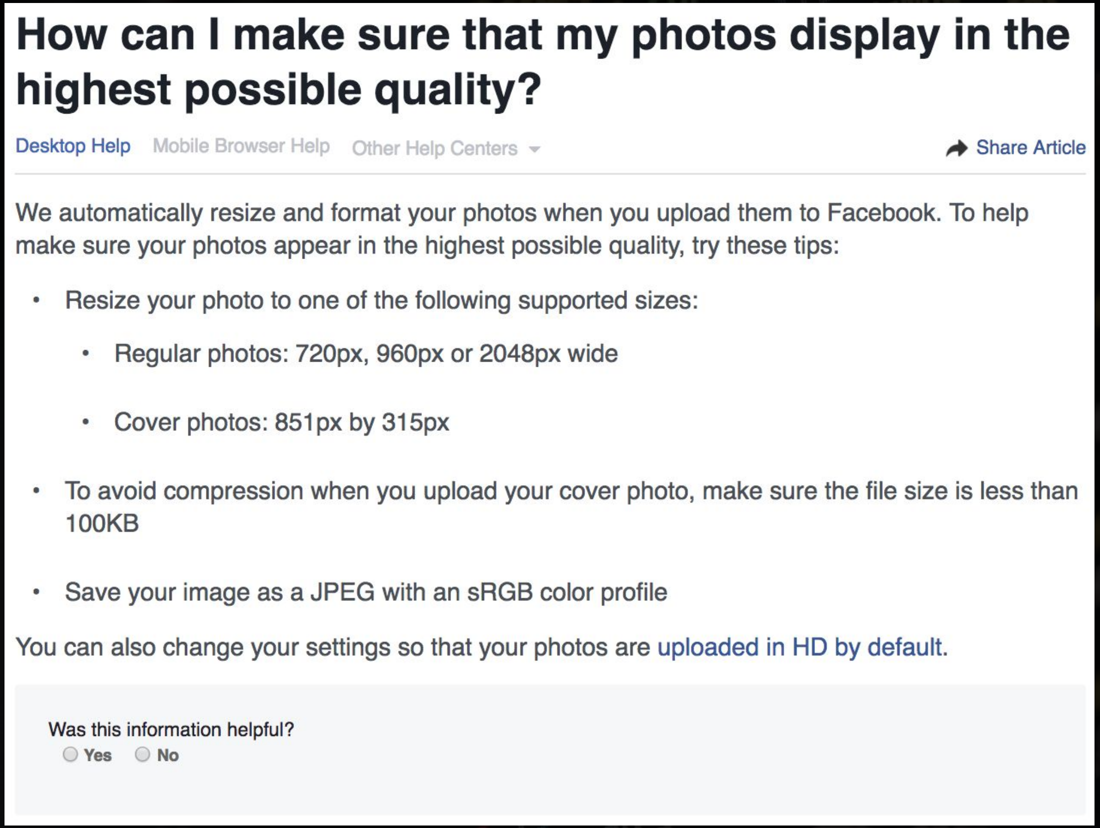
A steganography tool for hiding secret messages within photos on Facebook.

√ Send secret messages on Facebook
√ Hide messages in photos
√ Protect your messages from being read by Facebook, Governments or prying friends.

Secretbook was built as a research project in Oxford University.

- Open-source Social Stego tool
- Chrome Extension (2013)
- Reverse engineered Facebook's lossy compression algorithm
- Allowed for payloads of up to 140 characters in length

Jamming Techniques



How can I make sure that my photos display in the highest possible quality?

We automatically resize and format your photos when you upload them to Facebook. To help make sure your photos appear in the highest possible quality, try these tips:

- Resize your photo to one of the following supported sizes:
  - Regular photos: 720px, 960px or 2048px wide
  - Cover photos: 851px by 315px
- To avoid compression when you upload your cover photo, make sure the file size is less than 100KB
- Save your image as a JPEG with an sRGB color profile

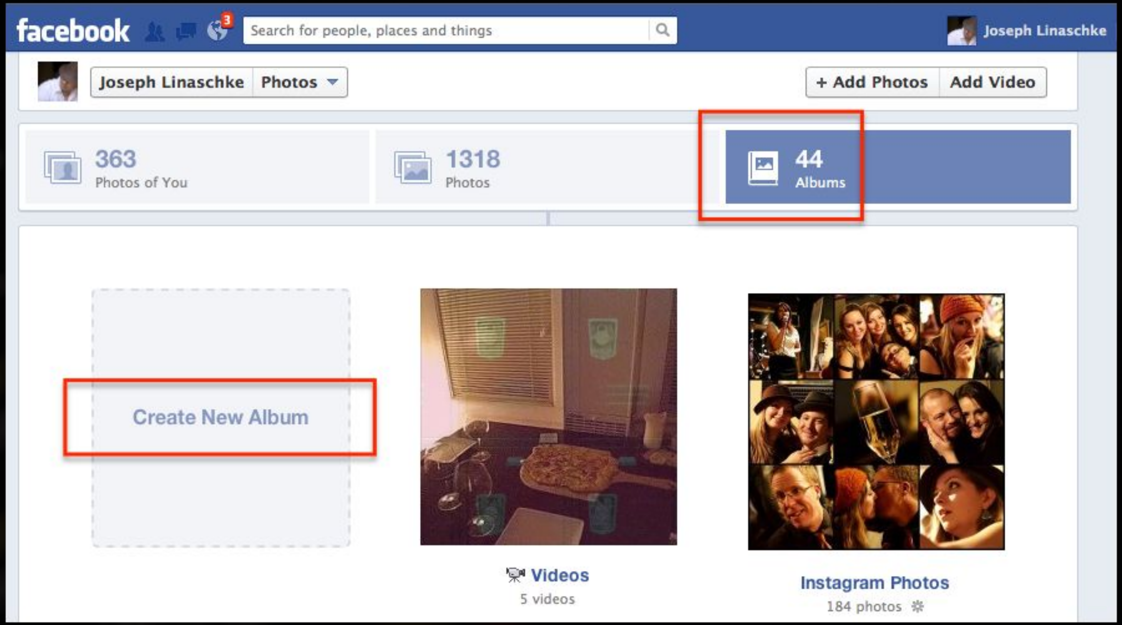
You can also change your settings so that your photos are uploaded in HD by default.

Was this information helpful?
○ Yes ○ No

- Server-side image upload restrictions and alterations
- Also legal concerns
  - Crime investigations
  - Trademark infringement
- Common Image upload Alterations:
  - Recompression
  - Metadata stripping
  - Filetype conversion
  - Resizing

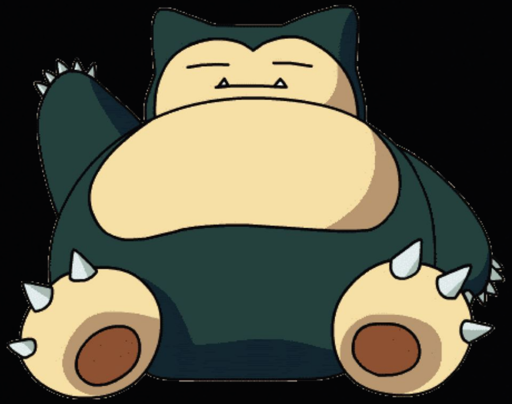
Bulk Image Uploads/Downloads

- Data Acquisition made easy
  - Permissive APIs for content creation
  - More content=more engagement=profit
- Off-the-shelf photo aggregators
  - Facebook albums
  - Pinterest boards
  - Flickr sets
  - Google+ Collections
- Or we can do it the 'hard way'
  - for photo in album{
    upload(photo); sleep(randInt); }



Auto-Generating Data

- Select 50k ImageNet samples
- Automate uploads and downloads
- =100k pre-uploaded and downloaded images
- Compare pixels between phases
- Can comparison/location be automated?
- But Neural Nets don’t scale to Images
    - width * height * 3 channels = unmanageable # weights
    - encode these properties into the architecture



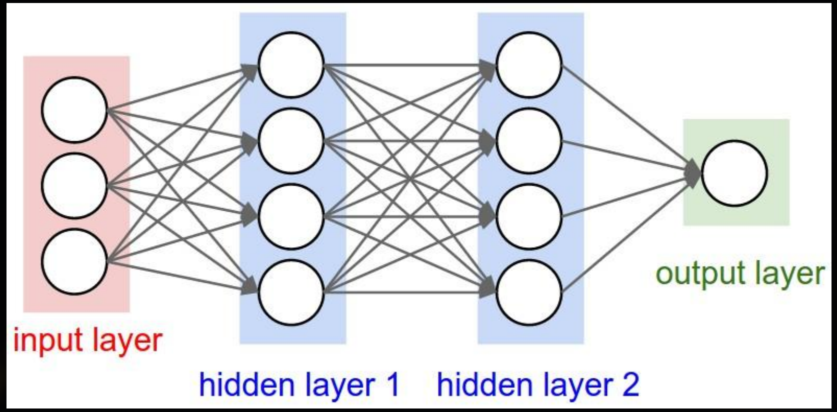
What humans see

```
08 02 22 97 38 15 00 75 04 05 07 78 52
49 49 99 40 17 81 18 57 60 87 17 40 98
81 49 31 73 55 79 14 29 93 71 40 67 53
52 70 95 23 04 60 11 42 69 24 65 56 54
22 31 16 71 51 67 63 89 41 92 36 54 22
24 47 32 60 99 03 45 02 44 75 33 53 78
32 98 01 20 64 23 67 10 26 38 40 67 59
67 26 20 68 02 62 12 20 95 63 94 39 63
```

What computers see

Convolutional Neural Networks

- Proven great for Computer Vision Tasks:
  - Object classification, Facial recognition
- Pose as a Regression Task
  - Locate optimally embeddable pixels
  - Akin to bounding boxes for object detection
- ConvNet hyperparameters
  - 7 stacked layers (5 convolutional, 2 fully connected)
  - Fed thru ReLUs and smooth L1 loss regression layer
- Model spec
  - Keras on top of TensorFlow (Python)
  - Google GPU (8 vCPU Nvidia Tesla)



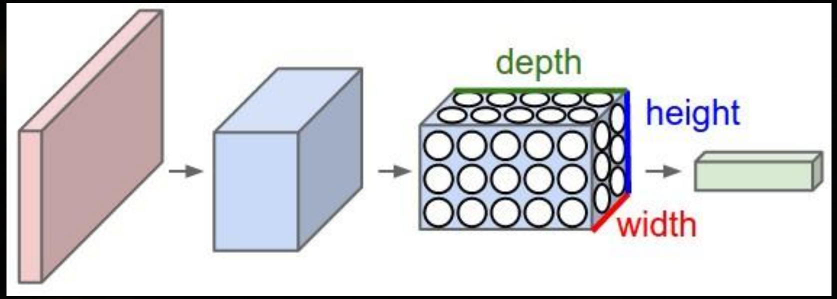


Illustration: Andrej Karpathy
CNNs: Szegedy, Toshev & Erhan, 2013

Prototype Evaluation

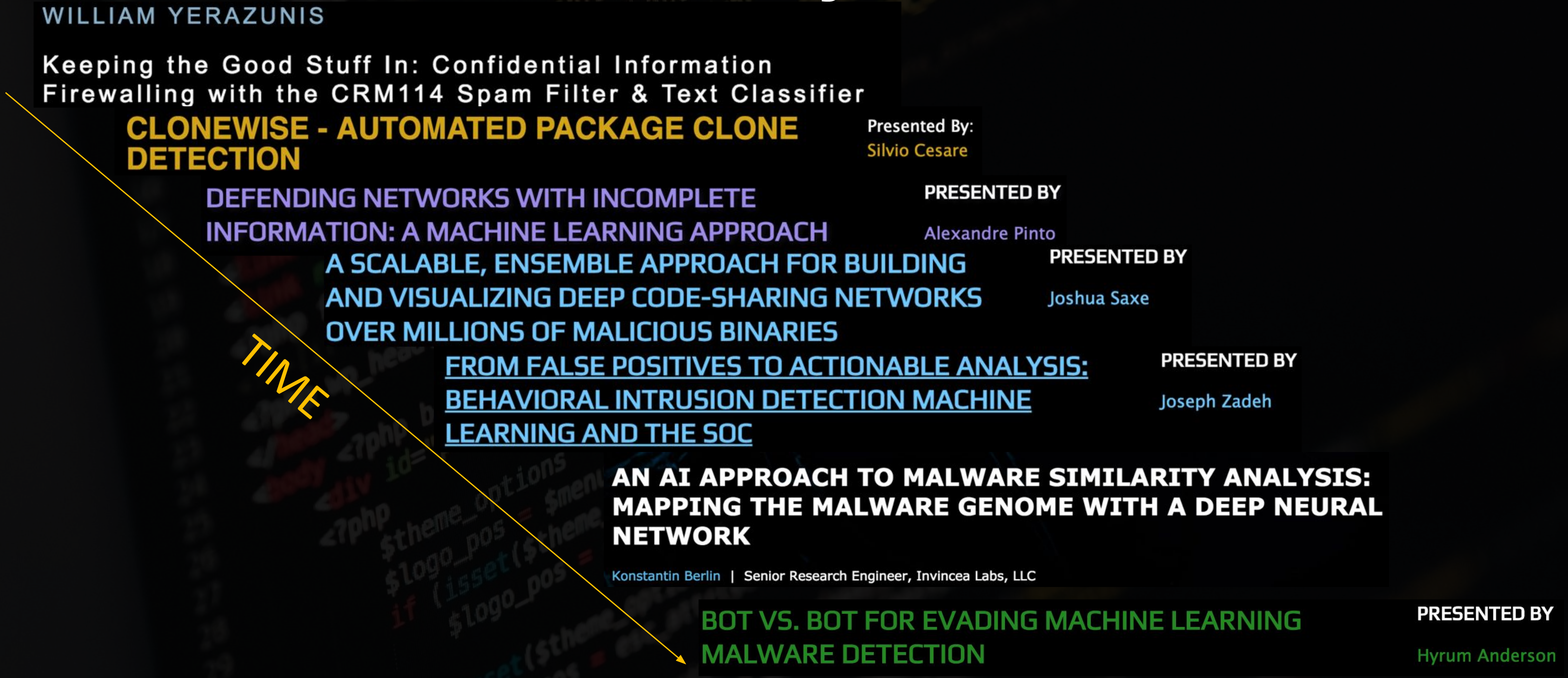
- More robust, less detectable transmission
- Learned locations correspond to locations that are more complex and “busier”
- Minimal Visual Dissimilarity
  - Distortion: peak signal-to-noise ratio
  - Capacity: byte Survivability
- Recovery rates worsen as hidden data size decreases

Data-Driven Red and Blue Teaming

A Picture is Worth A Thousand Words:

Deep Neural Networks for Social Stego

InfoSec ML Historically Prioritizes Defense



WILLIAM YERAZUNIS

Keeping the Good Stuff In: Confidential Information Firewalling with the CRM114 Spam Filter & Text Classifier

CLONEWISE - AUTOMATED PACKAGE CLONE DETECTION

Presented By:
Silvio Cesare

DEFENDING NETWORKS WITH INCOMPLETE INFORMATION: A MACHINE LEARNING APPROACH

PRESENTED BY
Alexandre Pinto

A SCALABLE, ENSEMBLE APPROACH FOR BUILDING AND VISUALIZING DEEP CODE-SHARING NETWORKS OVER MILLIONS OF MALICIOUS BINARIES

PRESENTED BY
Joshua Saxe

FROM FALSE POSITIVES TO ACTIONABLE ANALYSIS: BEHAVIORAL INTRUSION DETECTION MACHINE LEARNING AND THE SOC

PRESENTED BY
Joseph Zadeh

AN AI APPROACH TO MALWARE SIMILARITY ANALYSIS: MAPPING THE MALWARE GENOME WITH A DEEP NEURAL NETWORK

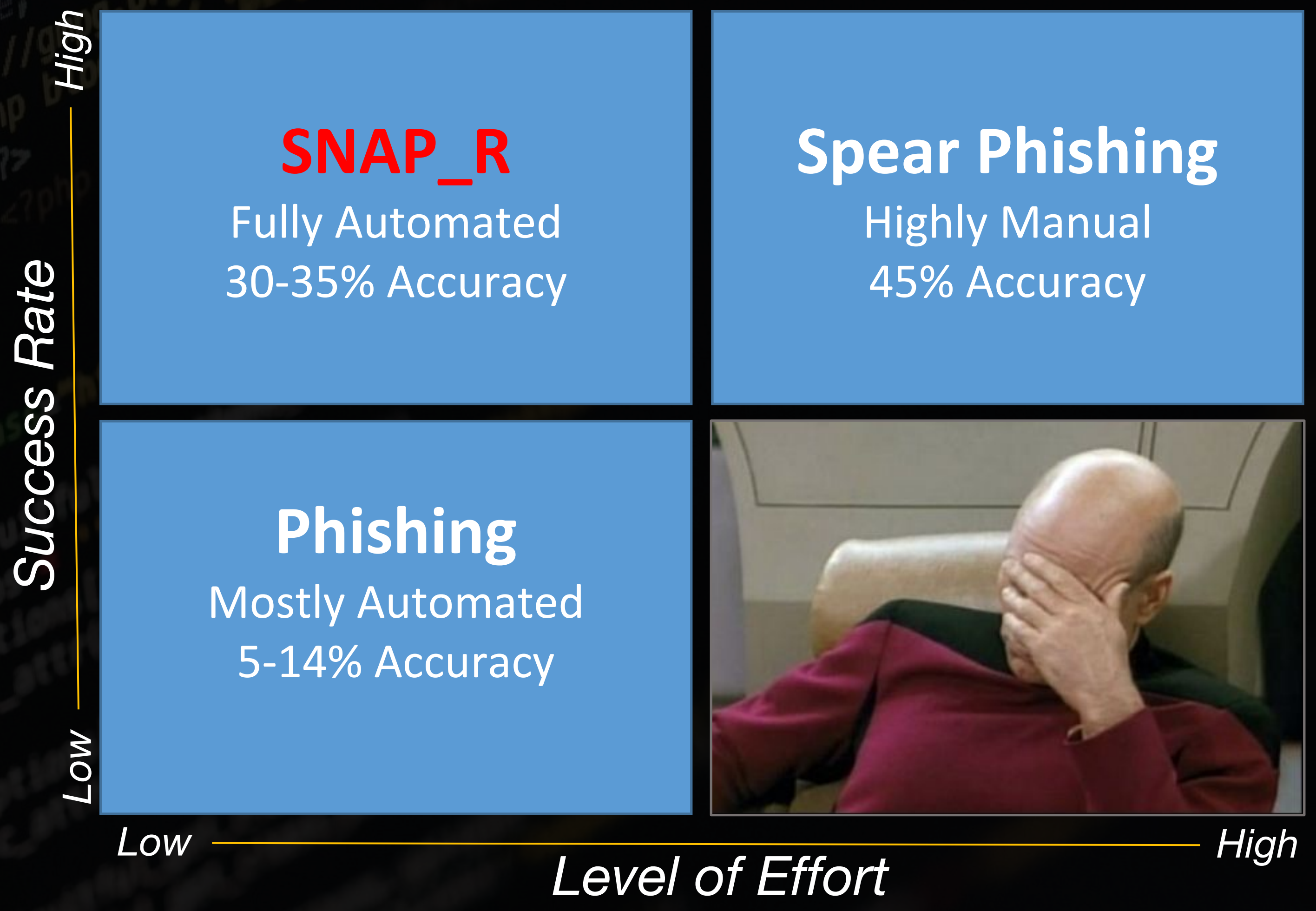
Konstantin Berlin | Senior Research Engineer, Invincea Labs, LLC

BOT VS. BOT FOR EVADING MACHINE LEARNING MALWARE DETECTION

PRESENTED BY
Hyrum Anderson

Data-Driven Social Engineering

- Black Hat/DEF CON 2016
- Why Twitter?
  - Bot-friendly API
  - Colloquial syntax
  - Shortened URLs
  - Abundant personal data
- Machine grammar suffices
- 10k+ DoD accts targeted



SNAP_R
Fully Automated
30-35% Accuracy

Spear Phishing
Highly Manual
45% Accuracy

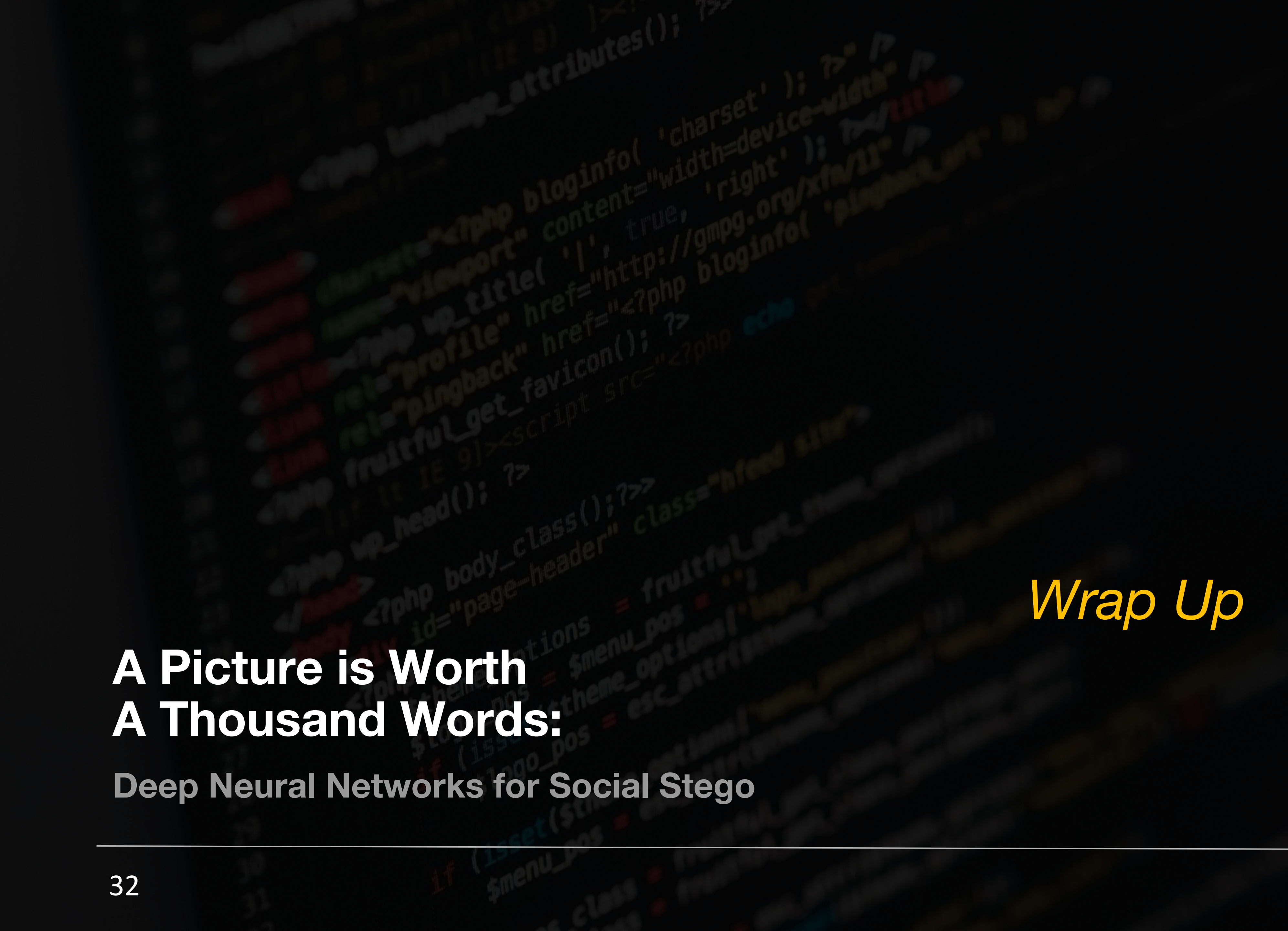
Phishing
Mostly Automated
5-14% Accuracy

Red Team ML Rising

- Growing number of examples:
  - Micro-targeted social engineering
  - Password cracking
  - Captcha subversion
  - AV evasion
  - Steganography
- Offensive ML easier than defensive ML!
  - “Labeling Bottleneck” - unsupervised
- Success matters more for blue than red team
- Retreating barriers to entry
  - More open-source initiatives
  - Cheapening access to powerful machines (eg. GPUs)

Not to worry, though...

- Offensive ML is a positive development
- It will “keep us honest”
- Emerging defenses keep pace:
  - Semi-supervised learning
  - Adversarial learning
  - Transfer learning
  - Self-supervised reinforcement learning
- Ultimately improve security
- Faster this is realized, the better

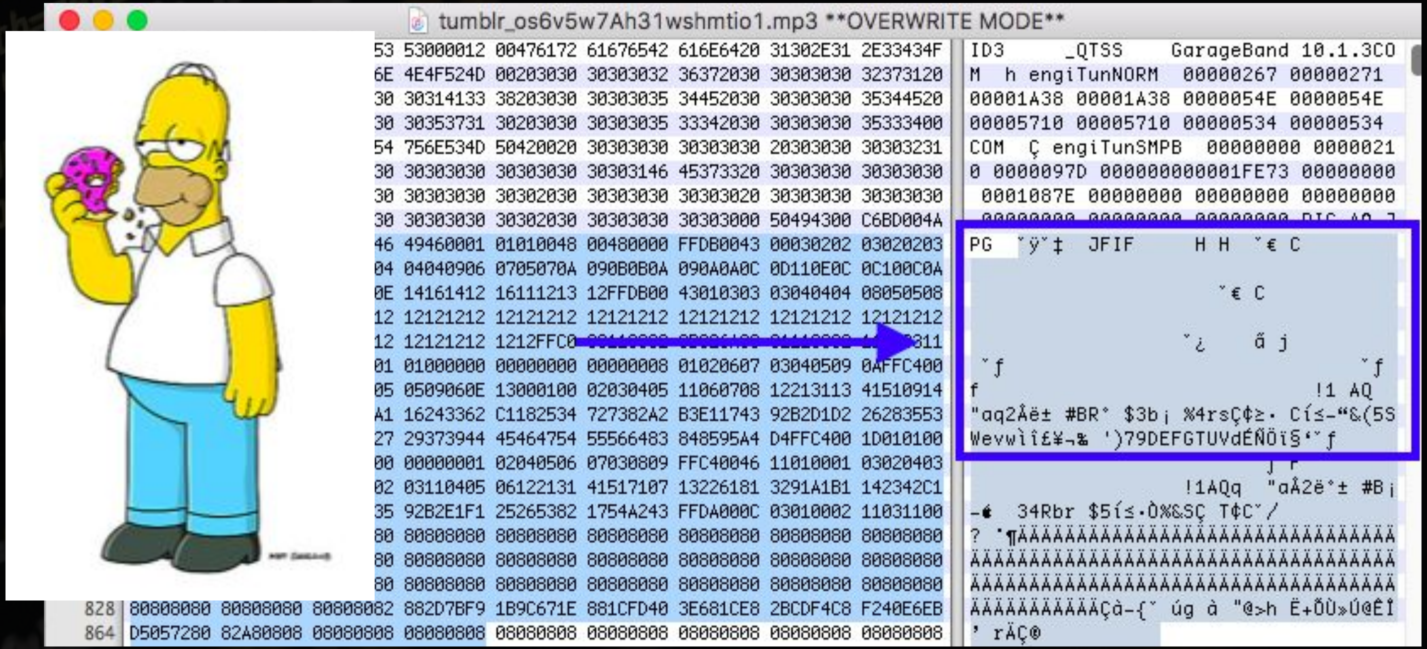


Wrap Up

A Picture is Worth A Thousand Words:

Deep Neural Networks for Social Stego

Next Steps



tumblr_os6v5w7Ah31wshmtio1.mp3 **OVERWRITE MODE**

```
53 53000012 00476172 61676542 616E6420 31302E31 2E33434F   ID3      _QTSS    GarageBand 10.1.3CO
6E 4E4F524D 00203030 30303032 36372030 30303030 32373120   M  h engiTunNORM  00000267 00000271
30 30314133 38203030 30303035 34452030 30303030 35344520   00001A38 00001A38 0000054E 0000054E
30 30353731 30203030 30303035 33342030 30303030 35333400   00005710 00005710 00000534 00000534
54 756E534D 50420020 30303030 30303030 20303030 30303231   COM  Ç engiTunSMPB  00000000 0000021
30 30303030 30303030 30303146 45373320 30303030 30303030   0 0000097D 00000000001FE73 00000000
30 30303030 30302030 30303030 30303020 30303030 30303030    0001087E 00000000 00000000 00000000
30 30303030 30302030 30303030 30303000 50494300 C6BD004A    00000000 00000000 00000000 PIC AQ J
46 49460001 01010048 00480000 FFDB0043 00030202 03020203   PG  ˇÿˇ‡  JFIF     H H  ˇ€ C
04 04040906 0705070A 090B0B0A 090A0A0C 0D110E0C 0C100C0A
0E 14161412 16111213 12FFDB00 43010303 03040404 08050508                         ˇ€ C
12 12121212 12121212 12121212 12121212 12121212 12121212
12 12121212 1212FFC0 [illegible] 311                        ˇ¿     ǻ j
01 01000000 00000000 00000008 01020607 03040509 0AFFC400    ˇƒ                              ˇƒ
05 0509060E 13000100 02030405 11060708 12213113 41510914   f                             !1 AQ
A1 16243362 C1182534 727382A2 B3E11743 92B2D1D2 26283553   "aq2Åë± #BR˚ $3b¡ %4rsÇ¢≥· Cí≤–“&(5S
27 29373944 45464754 55566483 848595A4 D4FFC400 1D010100   Wevwìî£¥¬‰ ')79DEFGTUVdÉÑÖï§‘ˇƒ
00 00000001 02040506 07030809 FFC40046 11010001 03020403                         j r
02 03110405 06122131 41517107 13226181 3291A1B1 142342C1                     !1AQq  "aÅ2ë˚± #B¡
35 92B2E1F1 25265382 1754A243 FFDA000C 03010002 11031100   –  34Rbr $5í≤·Ò%&SÇ T¢Cˇ/
80 80808080 80808080 80808080 80808080 80808080 80808080   ? ˙¶ÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄ
80 80808080 80808080 80808080 80808080 80808080 80808080   ÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄ
80 80808080 80808080 80808080 80808080 80808080 80808080   ÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄÄ
828 80808080 80808080 80808082 882D7BF9 1B9C671E 881CFD40 3E681CE8 2BCDF4C8 F240E6EB   ÄÄÄÄÄÄÄÄÄÄÄÇà–{ˇ úg à "@>h Ë+ÕÙ»Ú@ÊÎ
864 D5057280 82A80808 08080808 08080808 08080808 08080808 08080808 08080808 08080808   ’ rÄÇ®
```

- More social networks
- More stego (frequency domain)
- Video files (MP4, MOV, etc.)
  - Soon-to-be most popular
  - News Feed promoted
- Audio files (MP3)
  - Create custom MP3s w/ GarageBand
  - MP3s embedded JPEG insertion
  - ID3 Headers DC 24 SkyTalks Hosmer/Raggo www.python-forensics.org

Mitigations

- More sophisticated and dynamic jamming techniques
- Anomaly/Histogram analysis - increased quantization
- Impermanence: delete by default
  - Ephemeral images a la Snapchat
- But generally, steganalysis is hard!
  - Variance in social networks add exponential complexity to identifying existence of stego and recovery of evidence - “know thy enemy”

Summary and Questions?

Philip Tully **Mike Raggo**

@phtully @datahiding

ZEROFOX®

802secure

- Social networks and image hosting services can be orthogonally used to transmit data covertly
- Steganography can be automated despite distorting image upload side effects
- Offensive AI is cheaper and easier to implement than defensive AI
- Code released soon, PRs welcome