A Picture is Worth a Thousand Words, Literally:
Deep Neural Networks for Social Stego

Philip Tully | Mike Raggo
#whoami

Philip Tully
@phtully
Principal Data Scientist at ZeroFOX
PhD (KTH & University of Edinburgh)
Machine Learning and Neural Nets

Mike Raggo
@datahiding
CSO @802 Secure, 17 yrs Stego Research
StegSpy DC12, Author “Data Hiding”
NSA National Cryptologic Museum
DC25: Community, Discovery and the Unintended Uses of Technology
A Picture is Worth a Thousand Words, Literally:

Deep Neural Networks for Social Stego

DIY Social Steganography

The Evolution of Steganography

Data-Driven Red and Blue Teaming

Wrap Up
The Evolution of Steganography

A Picture is Worth a Thousand Words, Literally:
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.”

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
- VOIP, RTP and UDP techniques emerge along with decoys to complicate detection and recovery
- Multimedia begins to evolve into a viable method
- Social Media Stego used in images and URLs with malware, CnC, and covert communications

Evolution Chart:
- Tool Count
- Dates: Sep-01, Jan-03, Sep-05, Aug-09, Oct-11
- Tool Counts: 88, 129, 451, 525, 742, 754, 757

WetStone Labs Collected Steganography Programs Since January 1999 Includes versions
Evolution of Stego - Internet Era

- Stego Apps Decoy Techniques (OpenPuff)
- Stealth Alternate Data Streams (NT)
- Weaponized CnC - Operation Shady RAT
- Protocols - VOIP, RTP, UDP => WiFi StegoStuffing, Bluetooth (Hosmer/Raggo - Wall of Sheep/Skytalks DEF CON 23 & 24)
- SmartWatch SWATtackhide.py Tizen SDK - Mike Raggo - DEF CON 23 Demo Labs & HackCon
- MP3 ID3 Metadata exploitation - Hosmer/Raggo Skytalks DC24
Types of Steganography

- Text/Linguistic Stego - 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.
A Picture is Worth a Thousand Words, Literally:

Deep Neural Networks for Social Stego
Social Network Photo Targets

- Profile Image
- Background Image
- Posted Image(s)
- Photo albums
- DM images
- Links to images on other websites
Carrier Image File Types

- **Image quality properties:**
  - Lossy v. Lossless Raster Compression

- **Common file formats:**
  - JPEG (Lossy)
  - PNG (Lossless)
  - TIFF (Lossless)
  - GIF (Lossless)
  - BMP (Lossless)
Trial and Error - Attempted Methods

- Metadata fields (varies by image types JPEG EXIF vs. JFIF, etc.)
- LSB - Least Significant Bit
- Insertion
- Append after EOF marker
- Linguistic Steganography
- Round trip: pre/post upload

```
R = 1101101X
G = 1001011X
B = 1001010X
```

DataGenetics
High-Level Testing Workflow

1. upload
2. download
3. diff()
<table>
<thead>
<tr>
<th>Social Network</th>
<th>Profile Photo</th>
<th>Post an Image</th>
<th>Background Image</th>
<th>Album, Book, Board</th>
<th>Round-trip (pre/post upload)</th>
<th>Audio (MP3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinterest</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insertion</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSB</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instagram</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Twitter</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Facebook</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Slack</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Insertion</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>LSB</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Tumblr</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Insertion</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>LSB</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Metadata</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Google+</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Insertion</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>LSB</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
A Picture is Worth a Thousand Words, Literally:

Deep Neural Networks for Social Stego
Signals in the Social Noise

- 4.75 billion pieces of content shared per day.
- 100+ hours of video uploaded per minute.
- 80+ million images uploaded per day.
- 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 shared
- Social posts containing images produce 650% higher engagement than text alone
Social Networks as Stego Conduits

- Heavily trafficked, tons of images
- Public nature and broadcast capabilities
- Convenient APIs for sharing (uploading / downloading) content for devs & apps
- Fake account creation is trivial
- Lack of IoC’s from network perspective
- Wild examples - C&C, malware, phishing
Social Stego in the Wild

Black Hat: HAMMERTOSS [FireEye]

White Hat: Instegogram [ENDGAME]
Secretbook by Owen Campbell-Moore

- 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
- Other heuristic DCT schemes exist
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); }

```cpp
if (isset($Int);
$param_exists = true;
$variable = $stored_value;

if ($variable != null) {
  // do something
}
```
Automated High-Level Testing Workflow

1. upload
2. download
3. diff()
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.

- Server-side image upload restrictions and alterations
  - Fast mobile content delivery

- Common Image upload Alterations:
  - Compression
  - Lowpass filtering (slight blur)
  - Metadata stripping
  - Filetype conversion
  - Resizing
  - Alpha compositing
Targeting Unaltered Carrier Pixels
Auto-Generating Data

- Select ~50k samples (e.g. ImageNet)
- Automate uploads and downloads
- =100k pre-uploaded and downloaded images
- Compare pixels between phases
- Can location choices be automated?
- ‘Classic’ Neural Nets don’t scale to images
  - width * height * 3 channels = unmanageable # weights
  - encode these properties into the architecture
Convolutional Neural Networks

- Proven great for Computer Vision Tasks
  - Object classification, Facial recognition

- Pose as Binary Classification Tasks
  - Locate optimally embeddable pixels
  - Akin to image segmentation
  - Feedforward networks and function approximation

- Model spec
  - Keras on top of TensorFlow (Python)
  - Google GPU (8 vCPU Nvidia Tesla)
  - Contracting/expanding, ~23 layers fed thru ReLUs

Illustration: Andrej Karpathy

CNNs: Szegedy, Toshev & Erhan, 2013
Image Segmentation - Predict Binary Masks

Left: DeepMask [Facebook Research]  
Above: u-net [Ronneberger et al]
Prototype Evaluation

- More robust, less detectable transmission
- Recovery rates worsen with len(hidden data)
  - 94.1% accuracy (FPs=lost data, FNs=lower capacity)
- Minimizes Visual Dissimilarity
  - Distortion: peak signal-to-noise ratio, MS-SSIM
  - Capacity: bit survivability
  - Otherwise, watermarking
- Learned pixels correlate w/ carrier locations that are more complex and “busier”
Innovation and Novelty

Spatial stego = more storage capacity than frequency stego, compression-intolerant

Previous ad hoc approaches weren’t data-driven. Learn from uploads (feedback)

- Updated processing logic = retrain
- In principle, generalizes across social networks

No need to know implementation details of compression or other nonlinear processing

- Documentation not usually available anyway
Data-Driven Red and Blue Teaming

A Picture is Worth a Thousand Words, Literally:

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

DEFENDING NETWORKS WITH INCOMPLETE INFORMATION: A MACHINE LEARNING APPROACH

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

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

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

Konstantin Berlin | Senior Research Engineer, Invincea Labs, LLC
Data-Driven Social Engineering

- DEF CON 24
- Why Twitter?
  - Bot-friendly API
  - Colloquial syntax
  - Shortened URLs
  - Abundant personal data
- Machine grammar suffices

![Diagram with three categories of social engineering strategies: 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 a positive development
- It will “keep us honest”
- Emerging defenses keep pace:
  - Semi-supervised learning
  - Adversarial learning
  - Transfer learning
  - Self-supervised reinforcement learning
- Ultimately fortify security
- Faster this is realized, the better
A Picture is Worth a Thousand Words, Literally:
Deep Neural Networks for Social Stego
• Data exfiltration, digital dead drops, C&C
• Bypass online censors
• Privacy - Metadata tracks thru social media. Strip it if there’s concern

• Piracy - copyright in metadata
• Social media security awareness
Next Steps

- More social networks, crypto
- Deal w/ filters, resizing
- Fragment/Disperse payload
- Test more file types
  - Video files (MP4, MOV, etc.)
    - News Feed promoted, soon-to-be most popular
  - Audio files (MP3)
    - Create custom MP3s w/ GarageBand, embedded JPEG insertion
    - ID3 Headers DC 24 SkyTalks Hosmer/Raggo

www.python-forensics.org
Mitigations

- More dynamic jamming techniques
- Histogram “zigzag” - color quantization
  - Statistical: Means, variances, chi-square tests, linear analysis, wavelet statistics, kurtosis
- Impermanence: delete by default
  - Ephemeral images a la Snapchat
- Steganalysis is hard w/o access to orig image
  - Further obscurement through social’s scale, variance
Summary and Questions

- 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 to be released on GitHub piecemeal, followed by technical report (WIP)