Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

John Seymour  |  Philip Tully
You care about phishing on social media
TL;DR

#SNAP_R

Social Network Automated Phishing with Reconnaissance
ISO: Demo Volunteers

Tweet #SNAP_R before the demo to get an example tweet!
<table>
<thead>
<tr>
<th></th>
<th>John Seymour</th>
<th>Philip Tully</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whoami name</td>
<td>@_delta_zero</td>
<td>@phtully</td>
</tr>
<tr>
<td>Profession</td>
<td>Data Scientist at ZeroFOX</td>
<td>Senior Data Scientist at ZeroFOX</td>
</tr>
<tr>
<td>Education</td>
<td>Ph.D. student at UMBC</td>
<td>Ph.D. student at University of Edinburgh &amp; Royal Institute of Technology</td>
</tr>
<tr>
<td>Research Focus</td>
<td>Researches Malware Datasets</td>
<td>Brain Modeling &amp; Artificial Neural Nets</td>
</tr>
</tbody>
</table>
A Novel Phishing Campaign Design

<table>
<thead>
<tr>
<th>Level of Effort</th>
<th>Success Rate</th>
<th>Phishing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>5-14% Accuracy</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>Mostly Automated</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>High Manual</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>&gt;30% Accuracy</td>
</tr>
</tbody>
</table>

- **Our #SNAP_R**: Fully Automated, >30% Accuracy
- **Spear Phishing**: Highly Manual, 45% Accuracy
- **Phishing**: Mostly Automated, 5-14% Accuracy
Fooling Humans for 50 Years

1966: ELIZA Chatbot
- Joseph Weizenbaum, MIT
- Parsing & keyword replacement

2016: @TayandYou
- Microsoft AI
- Deep Neural Network
InfoSec ML Historically Prioritizes Defense
Weaponizing Data Science for Social Engineering:
Automated E2E Spear Phishing on Twitter
Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter
Why Twitter?

- Bot-friendly API
- Colloquial syntax
- Shortened links
- Trusting culture
- Incentivized data disclosure

Nikita @Niki7a · 1h
I'm doing random #FF's till #DEFCON. Starting with: @_sn0ww #skilled #social-engineer #bbwinner #OSINT #uber #Rad
Where Do the Phishers Live? Collecting Phishers Geographic Locations from Automated Honeypots

Robbie Gallagher

We’ve taken a novel approach to automating the determination of a phishers geographic location. With the help of Markov chains, we craft honeypot responses to phishers’ emails in an attempt to beat them at their own game. We’ll examine the underlying concepts, implementation of the system and reveal some results from our ongoing experiment.
Techniques, Tactics and Procedures

- Our ML Tool...
  - Shortens payload per unique user
  - Auto-tweets at irregular intervals
  - Triages users wrt value/engagement
  - Prepends tweets with @mention
  - Obey rate limits

- We added...
  - Post non-phishing posts
  - Build believable profile
Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter
Triage of High Value Targets on Twitter

- Accessible personal info
- Historical profile posts
- Heterogeneous data
- Text, images, urls, stats, dates
Extracting Features from GET users/lookup

- Engagement: following/followers
- #myFirstTweet
- Default settings
- Description content
- Account age
Clustering Predicts High Value Users

Eric Schmidt
Selecting the Best Clustering Model

- Many algorithms
- Many hyperparameters
- Max avg. score $\in [-1, \ldots, 1]$
- 0.5-0.7 reasonable structure
Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter
Recon and Footprinting for Profiling

- Compute histogram of tweet timings (binsize = 1 hour)
- Random minute within max hour to tweet
- Bag of Words on timeline tweets
- Select most commonly occurring non-stopword
- We seed the neural network with topics that the user frequently posts about
Leveraging Markov Models

- Popular for text generation: see /r/SubredditSimulator, InfosecTalk TitleBot
- Calculates pairwise frequency of tokens and uses that to generate new ones
- Based on transition probabilities
- Trained using most recent posts on the user’s timeline
Training a Recurrent Neural Network

- Hosted on Amazon EC2
- Trained on g2.2xlarge instance (65¢ per hour)
- Ubuntu (ami-c79b7eac)
- Training set > 2M tweets
- Took 5.5 days to train
- 3 layers, ~500 units/layer

Illustration: Chris Olah (@ch402)
LSTMs: Hochreiter & Schmidhuber, 1997
# Tradeoffs and Caveats

<table>
<thead>
<tr>
<th>Metric</th>
<th>LSTM</th>
<th>Markov Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Speed</td>
<td>Days</td>
<td>Seconds</td>
</tr>
<tr>
<td>Accuracy</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Availability</td>
<td>Public</td>
<td>Public</td>
</tr>
<tr>
<td>Size</td>
<td>Large</td>
<td>Small</td>
</tr>
</tbody>
</table>
| Caveats      | • Deeper representation of natural language, generalizes well  

• Retraining required for new languages  

• Overfits to each user, can create temporally irrelevant tweets  

• Performs poorly on users with few tweets
Language and Social Network Agnosticism

- Markov models only use content on user’s timeline, which means they can automatically generate content in other languages.

@8dot8 Nos alegra mucho informar por 3ra vez a como patrocinador de 8.8 Villanos goo.gl/dw4ure

- For neural nets, you’d only need to scrape data from the target language and retrain.

- Both of these methods can also be applied to other social networks.
Evaluation and Metrics

Weaponizing Data Science for Social Engineering:
Automated E2E Spear Phishing on Twitter
Here’s a malicious URL...

<table>
<thead>
<tr>
<th>URL:</th>
<th><a href="http://justfolio.cnminteractive.com/">http://justfolio.cnminteractive.com/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection ratio:</td>
<td>6 / 67</td>
</tr>
<tr>
<td>Analysis date:</td>
<td>2016-07-06 12:45:13 UTC ( 7 hours, 48 minutes ago )</td>
</tr>
</tbody>
</table>

- **Netcraft**: Malicious site
- **Opera**: Malicious site
- **Sophos**: Malicious site
- **CLEAN MX**: Phishing site
- **Fortinet**: Phishing site
- **Kaspersky**: Phishing site
And, apparently goo.gl lets us shorten it!
goo.gl also gives us analytics
Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

Results and Demo
Wild Testing #SNAP_R

Woody Allen and Iron & Wine at the weather for Sunday #WinterWeather #freezing #belowzero https://goo.gl/wn42Vh
Every Wednesday is our Gruffalo style WOODLAND ADVENTURE!! https://goo.gl/3s5PCJ
Bonne première mi temps de Cabaye en tout cas bonne nuit à tout le monde sauf aux portugais https://goo.gl/23C4A0
sus icons son muy lindos, ya amé todos los que me ha respondido una mención que la noche hace mil. https://goo.gl/1YK9yJ
#VeteranOfTheDay Check out Happy life with Ice Cream! Available for the next few days via https://goo.gl/sBMnZH
Happy birthday to my unit https://goo.gl/iwAiwD
Over 20 blocks with Pokemon Go! Grosse Pointe North High School https://goo.gl/dvlM9a
Let’s get out the vote for the acorns https://goo.gl/XoTtmO
Nominate yourself for the tip! https://goo.gl/yBF32k
Online news story about our Teen-Read-a-Thon at Hayters Gap library Saturday mornings at 9:30, https://goo.gl/RbxR7r
tell me how u gon call me rude when you don’t even have your own shit figured out https://goo.gl/YyloKM
Gif Pokemon #PokemonGo #PokemonGoPolska #Pokemon https://goo.gl/Ws15pZ
ahahaha no way?? Yeah I saw One Direction in concert https://goo.gl/9Gpf9r
Stay connected on Twitter and share engaging content with your followers? You don’t have to! Use https://goo.gl/e56sqj
When someone asks if you want to hang out in Eric’s basement https://goo.gl/nZVYhF
ya Allah,,, mimpi yang bukan aku ni.. Maissh.. Awak.. Byk btol yg awak follow guma ig sy ni https://goo.gl/2TsBCW
Go on over and back to work on Kaiju today. Monster powers are fun. https://goo.gl/3Cr0Gz
the streets of the ugliest shits I’ve seen https://goo.gl/lhTreo
#Bitcoin Is Not What You Think - #GoogleAlerts #datareach #cybersecurity #payments https://goo.gl/Vivelb
detta är verkliga ett skämt. Påväg hem från alla riktningar på stan… https://goo.gl/mD22P1
Oye Kya Ho Aap!! https://goo.gl/rrmkxh
Découvrir la puissance de l’univers Marvel. https://goo.gl/dcPD0t
Pilot Experiment

- Via #SNAP_R we sent 90 “phishing” posts out to people using #cat
  - After 2 hours, we had 17% clickthrough rate
  - After 2 days, we had between 30% and 66% clickthrough rate

- Inside the Data
  - goo.gl showed 27 clickthroughs (30%) came from a t.co referrer
  - Unknown referrers might be caused by bots
  - With unique locations, clickthrough rate may be as high as 66%
# Man vs. Machine 2 Hour Bake Off

<table>
<thead>
<tr>
<th>Metric</th>
<th>User</th>
<th>Person</th>
<th>SNAP_R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Targets</td>
<td>~200</td>
<td>819</td>
<td></td>
</tr>
<tr>
<td>Tweets/minute</td>
<td>1.67</td>
<td>6.85</td>
<td></td>
</tr>
<tr>
<td>Click-throughs</td>
<td>49</td>
<td>275</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>• Copy/Pasting messages to different hashtags</td>
<td>• Arbitrarily scalable with the number of machines</td>
<td></td>
</tr>
</tbody>
</table>
DEMO of #SNAP_R
Weaponizing Data Science for Social Engineering:
Automated E2E Spear Phishing on Twitter
Potential Use Cases

- Social media security awareness
- Social media security education
- Automated internal pentesting
- Social engagement
- Staff Recruiting
Mitigations

- Of course, we’re white hats here…
  - But machine learning is rapidly becoming automated, so black hats would have this capability soon.

- Protected accounts are immune to timeline scraping, which defeats the tool

- Bots can be detected

- Standard mitigations apply:
  - Don’t click on links from people you don’t know
  - Report! Twitter is pretty good at flagging spam accounts
  - Maybe URL shorteners should be responsible for malware?
Machine learning can be used offensively to automate spear phishing.

- Machine-generated grammar is bad, but Twitter users DGAF.
- Abundant personal data is publicly accessible and effective for social engineering.
We’ll also be at the **ZEROFOX** booth immediately after the presentation!

John Seymour  
@_delta_zero

Philip Tully  
@phtully