Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

John Seymour  |  Philip Tully

#WeaponizeMe
You care about phishing on social media

<table>
<thead>
<tr>
<th>Threat</th>
<th>Concern (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sophisticated attacks targeted directly at the organization</td>
<td>57%</td>
</tr>
<tr>
<td>Phishing, social network exploits or other forms of social engineering</td>
<td>46%</td>
</tr>
<tr>
<td>Accidental data leaks by end users who fail to follow security policy</td>
<td>21%</td>
</tr>
<tr>
<td>Polymorphic malware that evades signature-based defenses</td>
<td>20%</td>
</tr>
</tbody>
</table>
**TL;DR**

<table>
<thead>
<tr>
<th>Level of Effort</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Our Tool</strong></td>
<td>Fully Automated</td>
</tr>
<tr>
<td><strong>Spear Phishing</strong></td>
<td>Highly Manual</td>
</tr>
<tr>
<td><strong>Phishing</strong></td>
<td>Mostly Automated</td>
</tr>
</tbody>
</table>

Tweet **#WeaponizeMe** before the demo to get an example tweet!
# whoami

**John Seymour**  
@_delta_zero  
Data Scientist at ZeroFOX  
Ph.D. student at UMBC  
Researches Malware Datasets

**Philip Tully**  
@kingphish3r  
Senior Data Scientist at ZeroFOX  
Ph.D. student at University of Edinburgh & Royal Institute of Technology  
Brain Modeling and Artificial Neural Nets
Fooling Humans for 50 Years

1966: Eliza Chatabot
- 1966: ELIZA chatbot
- Joseph Weizenbaum, MIT

2016: @TayandYou AI Chatbot
- Microsoft
- Deep Neural Network

Eliza is a mock Rogerian psychotherapist. The original program was described by Joseph Weizenbaum in 1966. This implementation by Norbert Landsteiner 2005.
InfoSec ML Historically Prioritizes Defense

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

Presented By:
Silvio Cesare
Alexandre Pinto
Joshua Saxe
Joseph Zadeh

Konstantin Berlin | Senior Research Engineer, Invincea Labs, LLC

#WeaponizeMe
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 phisher’s 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 of the results
from our ongoing experiment.
Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter
Our #SMACK Tool

Twitter
Profiles
Payloads
Harvesting

#SMACK Tool

#WeaponizeMe
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
Techniques to Evade Detection

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

- We added...
  - Post non-phishing posts
  - Build believable profile
Design Flow

```
Design Flow

#SMACK Tool

Twitter Profiles Payloads Harvesting

is_target(user)

generate_timeline(depth)

gen_markov_tweet()

gen_nn_tweet()

schedule_tweet_and_sleep()

post_tweet_and_sleep()
```

#SMACK

Tool

Twitter Profiles Payloads Harvesting
Weaponizing Data Science for Social Engineering:
Automated E2E Spear Phishing on Twitter
Triage of High Value Targets on Twitter

Eric Schmidt
Triage of High Value Targets on Twitter

Eric Schmidt @ericschmidt · 10 Aug 2015
I think the Alphabet name is Awesome. googleblog.blogspot.com/2015/08/google…

Eric Schmidt @ericschmidt · 10 Aug 2015
Really excited about the vision and brilliance of Sundar.. he's going to be a great CEO! googleblog.blogspot.com/2015/08/google…

Eric Schmidt @ericschmidt · 5 Aug 2015
Feeling energized that US is embracing a clean power future. 2015 is shaping up to be a big year to #ActOnClimate.

Fact Sheet: President Obama to Announce Historic …
The Clean Power Plan is a Landmark Action to Protect Public Health, Reduce Energy Bills for Households and Businesses, Create American Jobs, and Bring
Extracting Features from GET users/lookup

- Engagement metrics
- #myFirstTweet
- Default settings
- Description content
- Account age

```json
{
    "follow_request_sent": false,
    "has_extended_profile": false,
    "profile_use_background_image": true,
    "profile_text_color": "333333",
    "default_profile_image": false,
    "id": 93957809,
    "profile_background_image_url_https": "https://pbs.twimg.com/profile_images/51150219793930240/3gfO0e77c_normal.jpeg",
    "verified": true,
    "profile_location": null,
    "profile_image_url_https": "https://pbs.twimg.com/profile_images/51150219793930240/3gfO0e77c_normal.jpeg",
    "profile_sidebar_fill_color": "DDDDDD",
    "listed_count": 26506,
    "status": {
        "contributors": null,
        "truncated": false,
        "text": "'Simplicity Is Complexity Resolved' - Constantin Brancusi",
        "in_reply_to_status_id": null,
        "id": 728409530782621137,
        "favorite_count": 577,
        "source": "<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client</a>\",
        "retweeted": false,
        "coordinates": null,
        "in_reply_to_screen_name": null,
        "in_reply_to_user_id": null,
        "in_reply_to_user_id_str": null,
        "id_str": "728409530782621137",
        "favorited": false,
        "geo": null,
        "in_reply_to_user_id_str": null,
        "lang": "en",
        "created_at": "Fri May 06 02:22:19 +0000 2016",
        "in_reply_to_status_id_str": null,
        "place": null
    },
    "is_translation_enabled": false,
    "utc_offset": -25200,
    "statuses_count": 356,
    "description": "Executive Chairman & former CEO",
    "friends_count": 235,
    "location": "Mountain View, CA",
    "profile_link_color": "0000FF",
    "profile_image_url": "https://pbs.twimg.com/profile_images/51150219793930240/3gfO0e77c_normal.jpeg",
    "following": false,
    "geo_enabled": false,
    "profile_banner_url": "https://pbs.twimg.com/profile_banners/93957809/1411004677",
    "profile_background_image_url": "https://pbs.twimg.com/profile_images/51150219793930240/3gfO0e77c_normal.jpeg",
    "name": "Eric Schmidt",
    "lang": "en",
    "profile_background_tile": false,
    "favourites_count": 0,
    "screen_name": "ericsschmidt",
}

16  #WeaponizeMe
Clustering predicts high value users
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
- Trained using the most recent posts on the user’s timeline

#SMACK Tool

Twitter
Profiles
Payloads
Harvesting
Inside LSTM Neural Networks

LSTM = Long Short Term Memory
Training the Neural Net

- Hosted on Amazon EC2
- Trained NN on g2.2xlarge instance (65¢ per hour)
- Ubuntu (ami-c79b7eac)
- Trained on more than 2 million tweets
- Took 3 days to train

LSTM = Long Short Term Memory
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.
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>

<table>
<thead>
<tr>
<th>Tool</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netcraft</td>
<td>Malicious site</td>
</tr>
<tr>
<td>Opera</td>
<td>Malicious site</td>
</tr>
<tr>
<td>Sophos</td>
<td>Malicious site</td>
</tr>
<tr>
<td>CLEAN MX</td>
<td>Phishing site</td>
</tr>
<tr>
<td>Fortinet</td>
<td>Phishing site</td>
</tr>
<tr>
<td>Kaspersky</td>
<td>Phishing site</td>
</tr>
</tbody>
</table>
And, apparently goo.gl lets us shorten it!

http://goo.gl/EJ09jc

http://justfolio.cnminteractive.com/
Created: 2016 Jun 16

Total Clicks 45

Clicks for the past: two hours | day | week | month | all time
goo.gl also gives us analytics
Initial Experiment

- 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%
Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter
Bake Off

We plan to do a humans vs. machine bake off prior to the conference. Stay tuned for the results!
Our #SMACK Tool In Action

Weaponizing Data Science for Social Engineering:
Automated E2E Spear Phishing on Twitter
DEMO of Our #SMACK Tool Framework
Weaponizing Data Science for Social Engineering:
Automated E2E Spear Phishing on Twitter
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.