Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

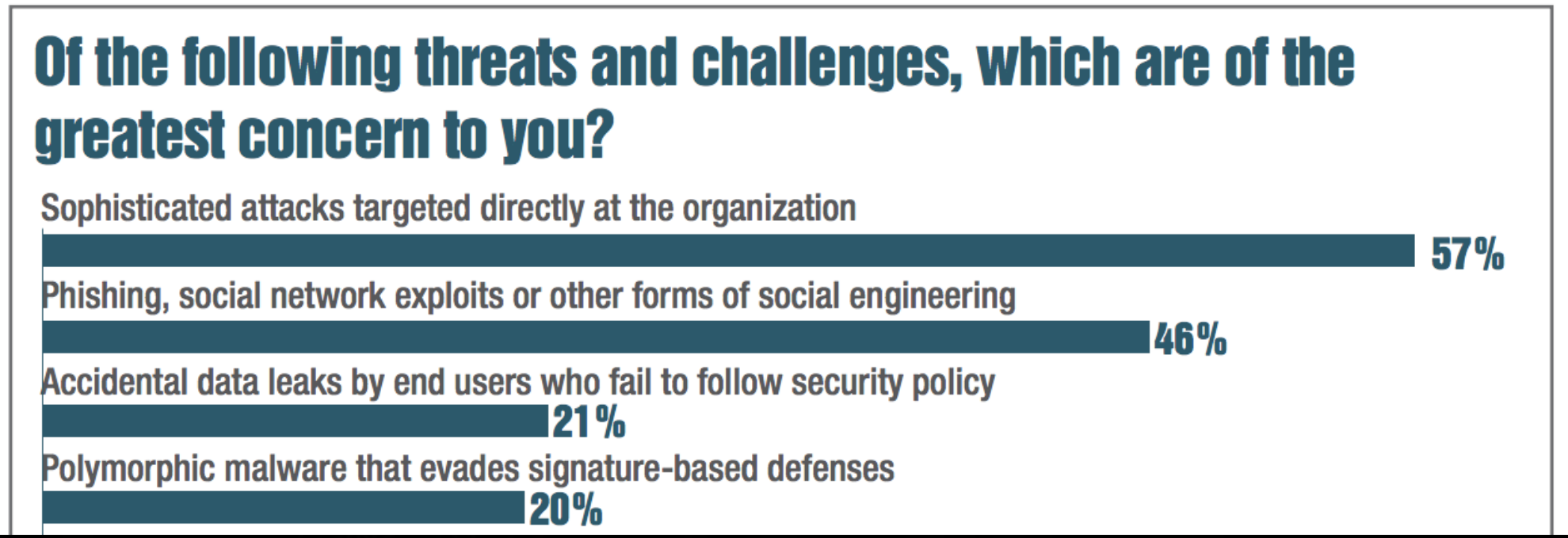
John Seymour | Philip Tully

#WeaponizeMe

You care about phishing on social media

Figure 1

The 2015 Black Hat Attendee Survey



#WeaponizeMe

TL;DR

Our Tool Fully Automated >30% Accuracy	**Spear Phishing** Highly Manual 45% Accuracy
Phishing Mostly Automated 5-14% Accuracy	

Suc[illegible]ate

Level of Effort

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

#whoami

John Seymour	Philip Tully
@_delta_zero	@kingphish3r
Data Scientist at ZeroFOX	Senior Data Scientist at ZeroFOX
Ph.D. student at UMBC	Ph.D. student at University of Edinburgh & Royal Institute of Technology
Researches Malware Datasets	Brain Modeling and Artificial Neural Nets

Fooling Humans for 50 Years

```
Welcome to
                EEEEEE  LL      IIII  ZZZZZZZ   AAAAA
                EE      LL       II       ZZ   AA   AA
                EEEEE   LL       II     ZZZ    AAAAAAA
                EE      LL       II    ZZ      AA   AA
                EEEEEE  LLLLLL  IIII  ZZZZZZZ  AA   AA

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


ELIZA: Please tell me what's been bothering you.
```

1966: Eliza Chatabot

- 1966: ELIZA chatbot
- Joseph Weizenbaum, MIT

20166: @TayandYou AI Chatbot

- Microsoft
- Deep Neural Network

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

TIME

#Shoutout

[Where Do the Phishers Live? Collecting Phishers' Geographic Locations from Automated Honeypots]

Robbie Gallagher

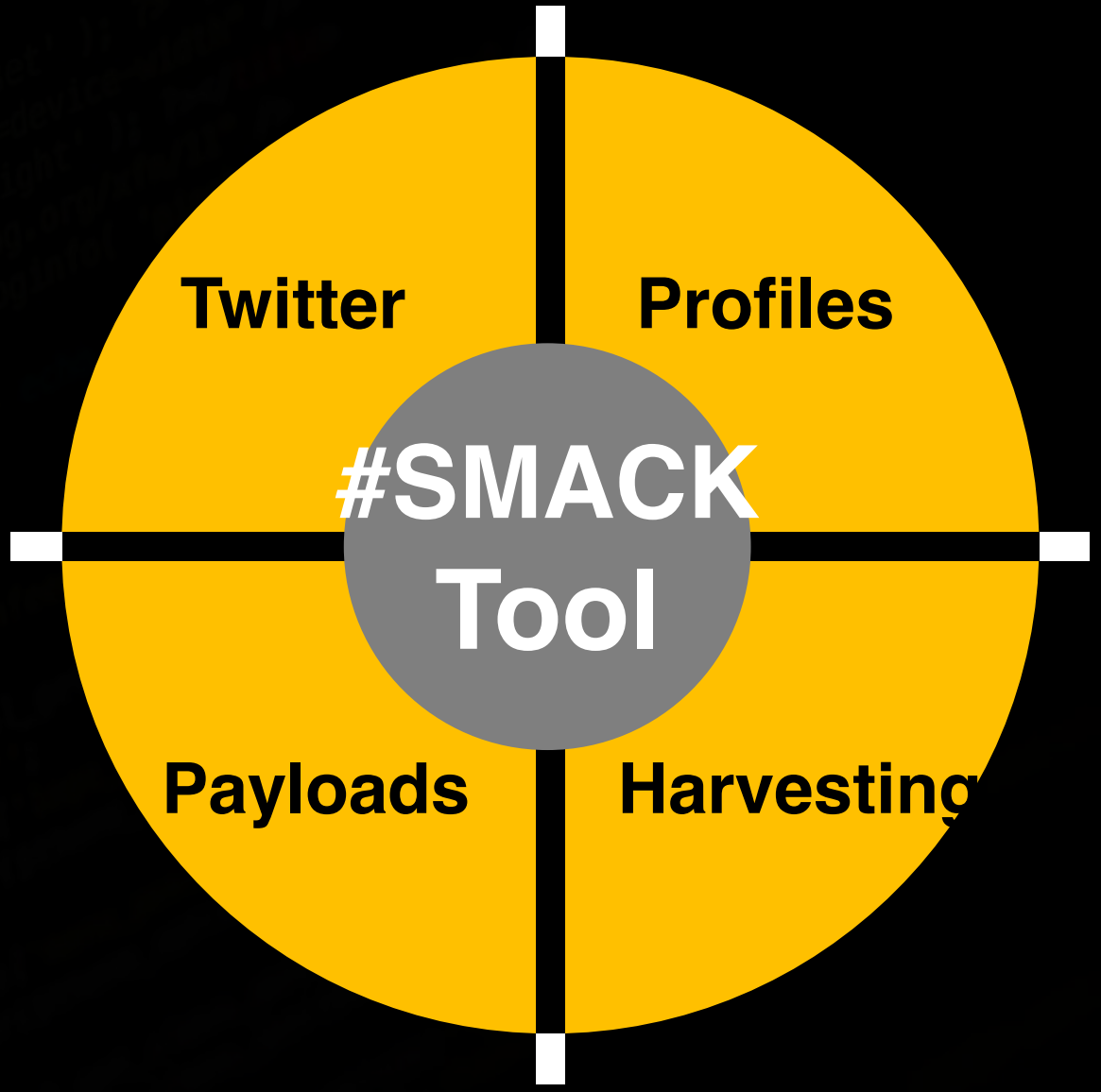
We've taken a novel approach to automating the determination of a phisher's geographic location. With the help of Markov chains, we craft honeypot responses to phishers' emails in an attempt to beat them at their own game. We'll examine the underlying concepts, implementation of the system, and reveal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

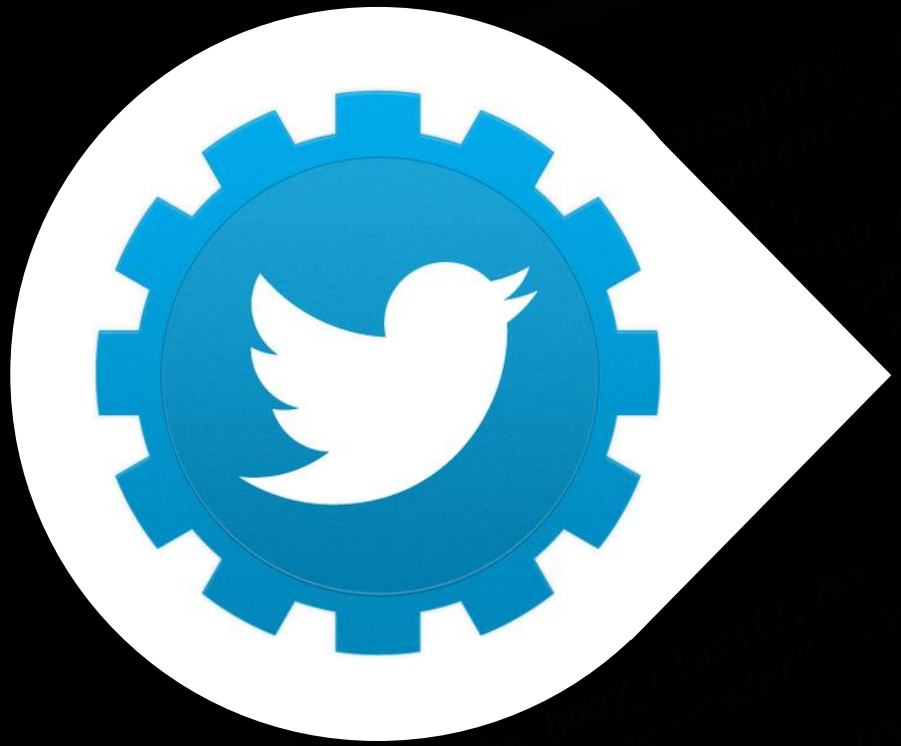
Machine Learning on Offense

Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

Our #SMACK Tool



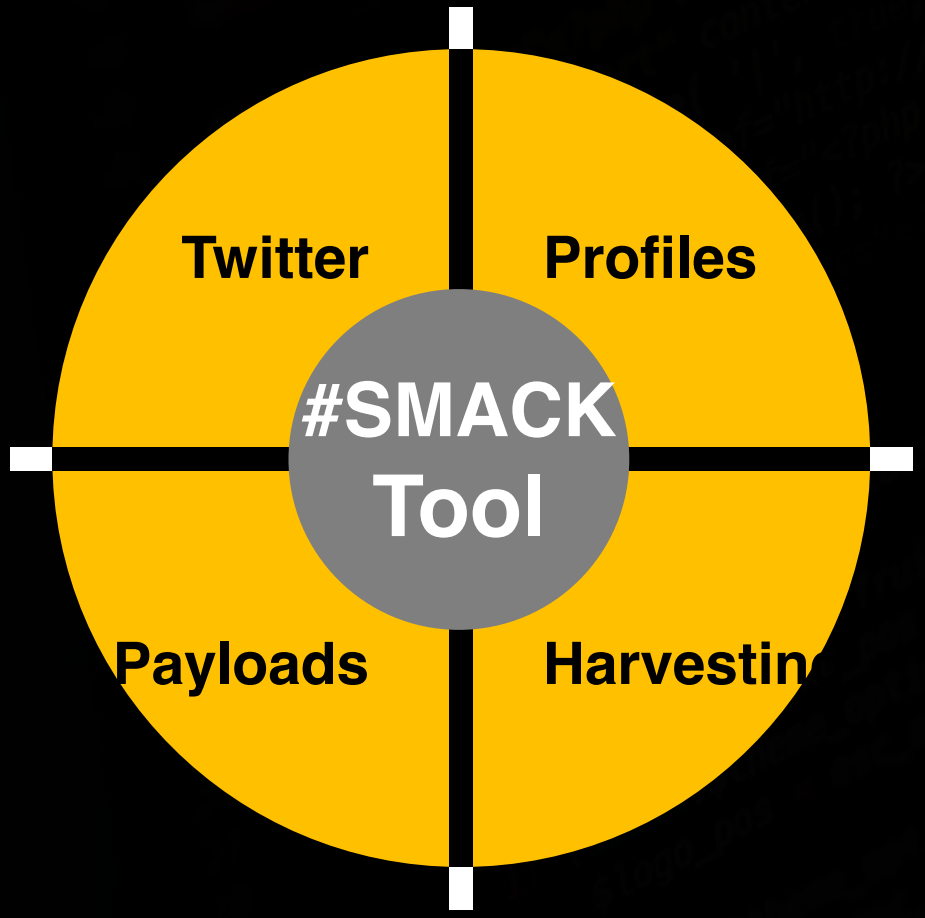


Why Twitter?

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

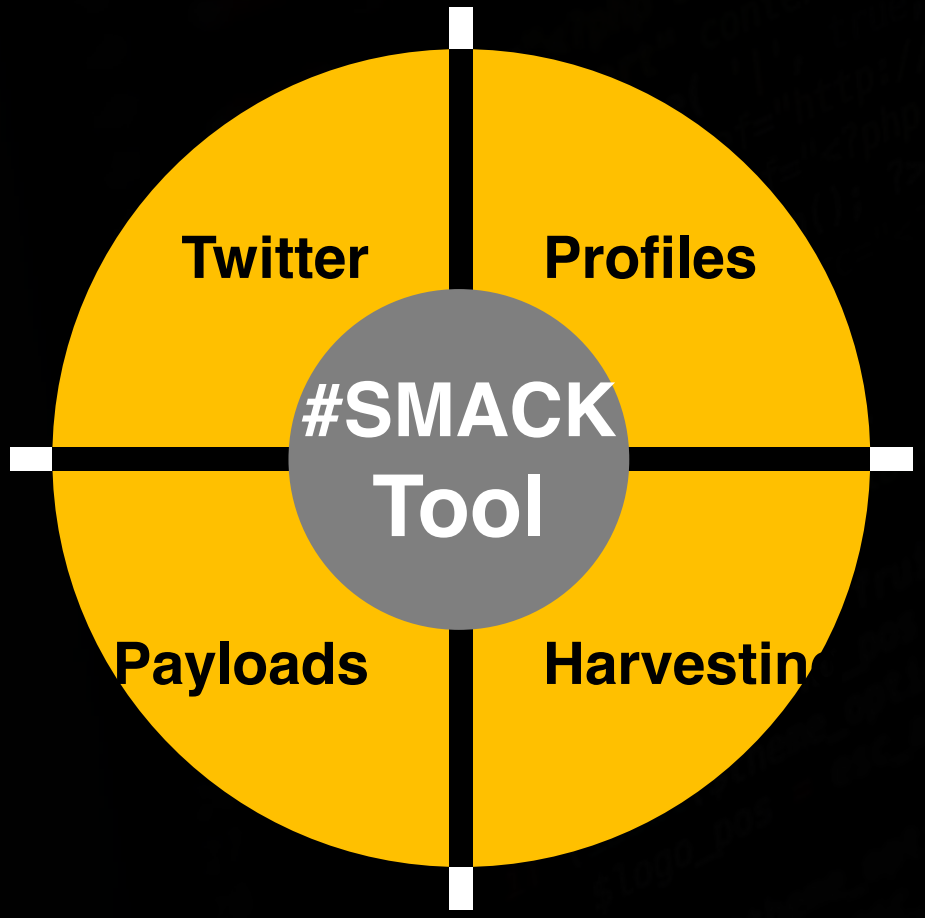
Nikita @Niki7a · 1h
I'm doing random #FF's till #DEFCON. Starting with: @_sn0ww #skilled #social-engineer #bbwinner #OSINT #uber #Rad

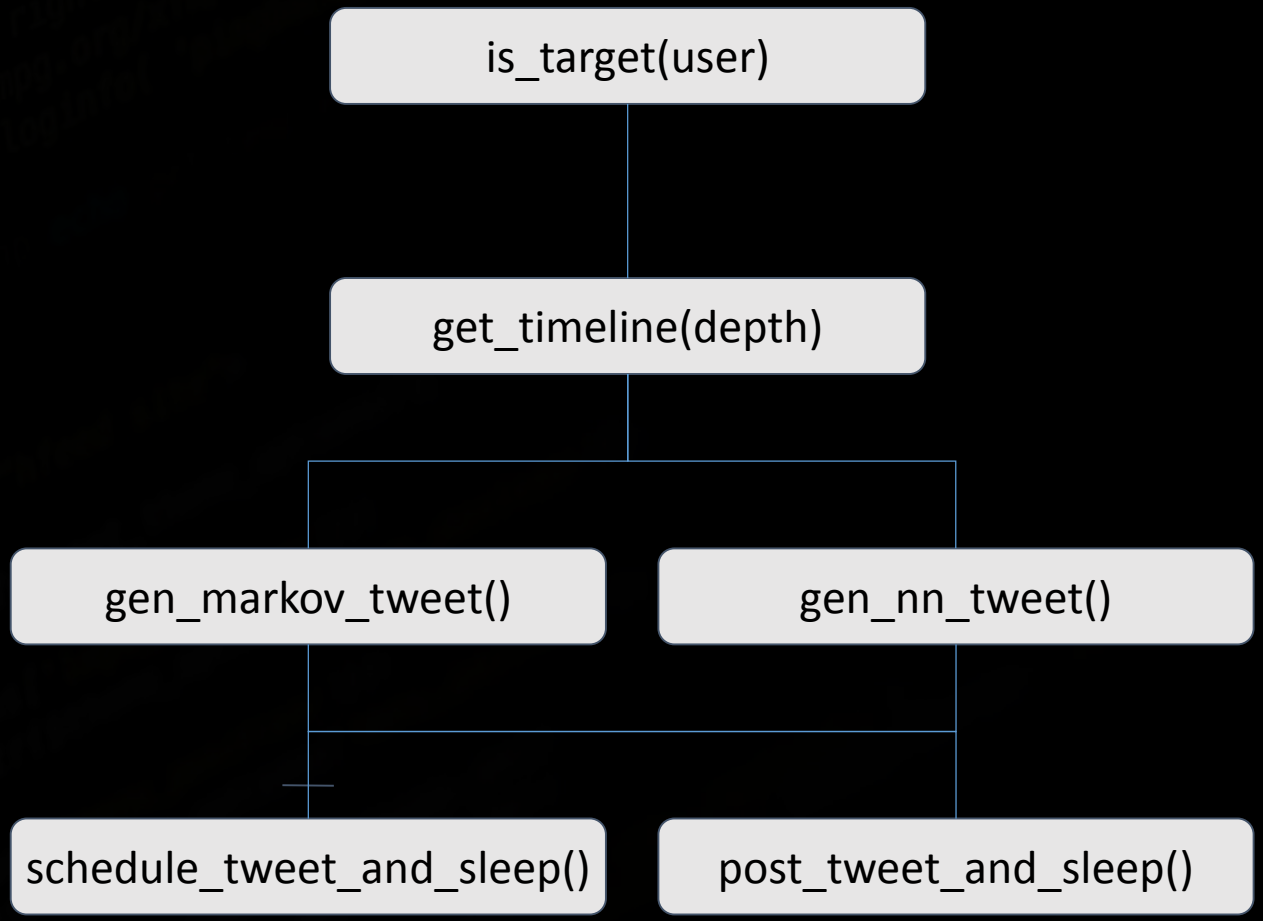
Techniques to Evade Detection



- 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


Twitter
Profiles
#SMACK
Tool
Payloads
Harvestin

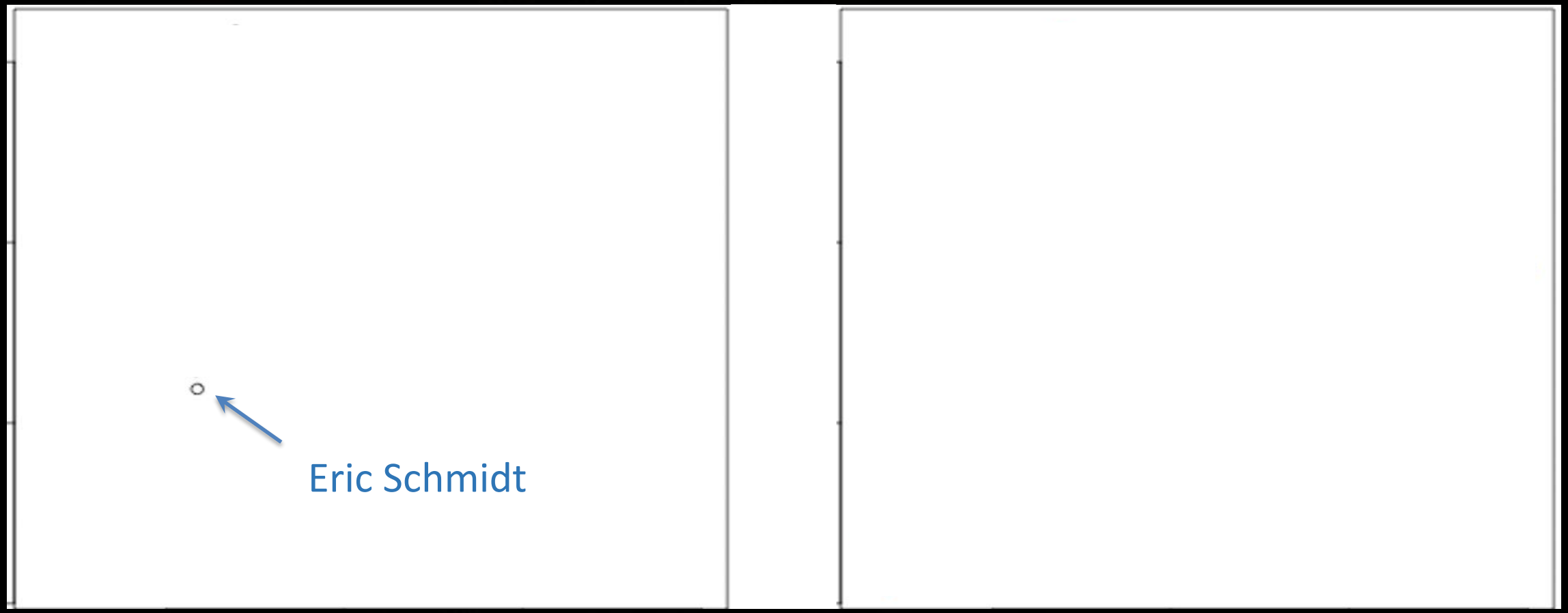

is_target(user)
get_timeline(depth)
gen_markov_tweet()
gen_nn_tweet()
schedule_tweet_and_sleep()
post_tweet_and_sleep()

Automated Target Discovery

Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

Triage of High Value Targets on Twitter



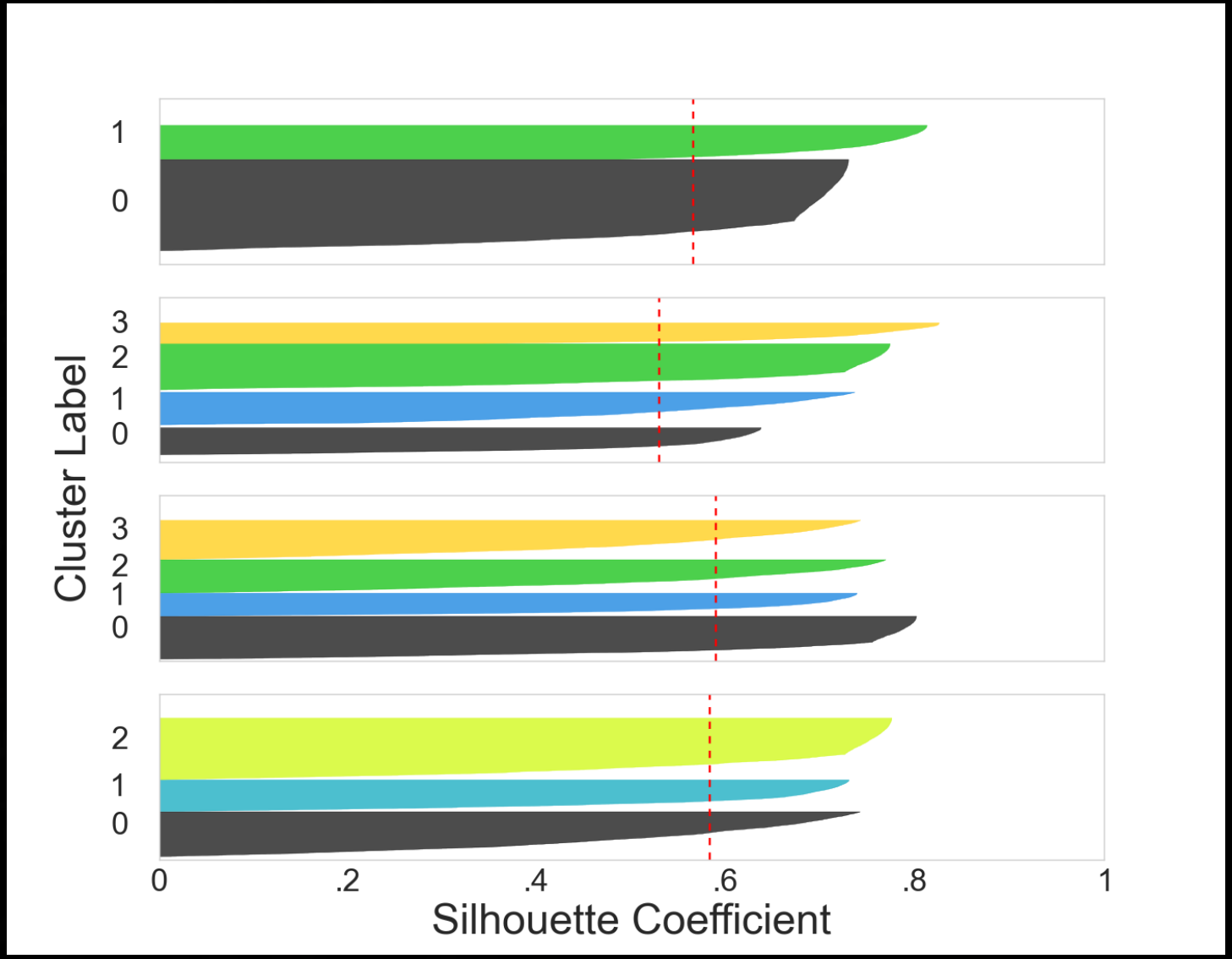
Triage of High Value Targets on Twitter

Extracting Features from GET users/lookup

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


```
{
  "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:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png",
  "verified": true,
  "profile_location": null,
  "profile_image_url_https": "https:\/\/pbs.twimg.com\/profile_images\/511582119793930240\/JgGOe77c_normal.jpeg",
  "profile_sidebar_fill_color": "DDEEF6",
  "listed_count": 20500,
  "status": {
    "contributors": null,
    "truncated": false,
    "text": "''Simplicity Is Complexity Resolved'' - Constantin Brancusi",
    "is_quote_status": false,
    "in_reply_to_status_id": null,
    "id": 7.2840953072021e+17,
    "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,
    "retweet_count": 450,
    "id_str": "728409530720210944",
    "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": 350,
  "description": "Executive Chairman & former CEO",
  "friends_count": 235,
  "location": "Mountain View, CA",
  "profile_link_color": "0084B4",
  "profile_image_url": "http:\/\/pbs.twimg.com\/profile_images\/511582119793930240\/JgGOe77c_normal.jpeg",
  "following": false,
  "geo_enabled": false,
  "profile_banner_url": "https:\/\/pbs.twimg.com\/profile_banners\/93957809\/1410806717",
  "profile_background_image_url": "http:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png",
  "name": "Eric Schmidt",
  "lang": "en",
  "profile_background_tile": false,
  "favourites_count": 0,
  "screen_name": "ericschmidt",
```

Clustering predicts high value users

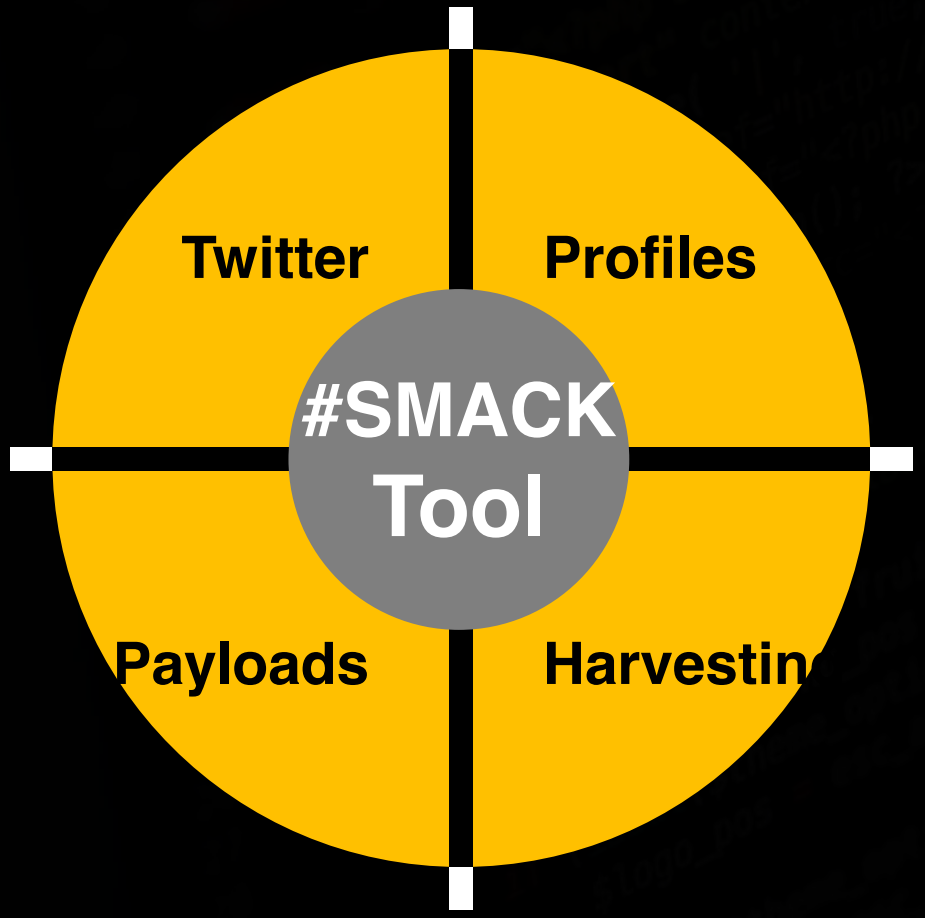


Automated Social Spear Phishing

Weaponizing Data Science for Social Engineering:

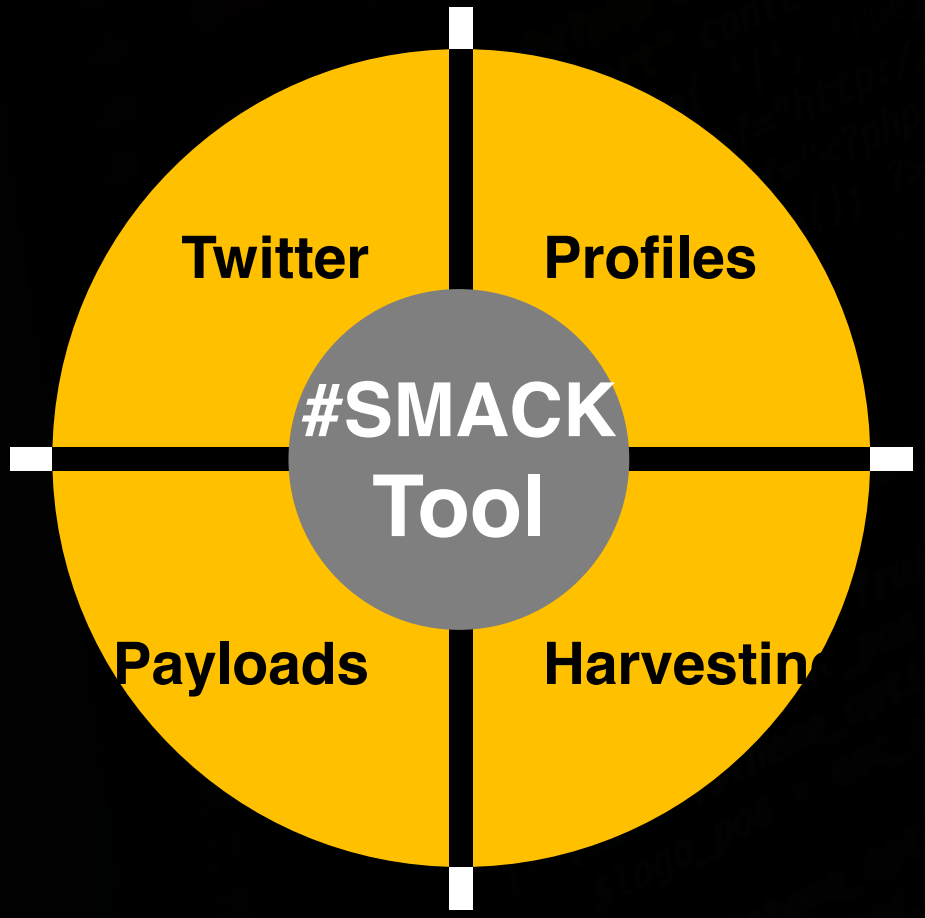
Automated E2E Spear Phishing on Twitter

Recon and Footprinting for Profiling



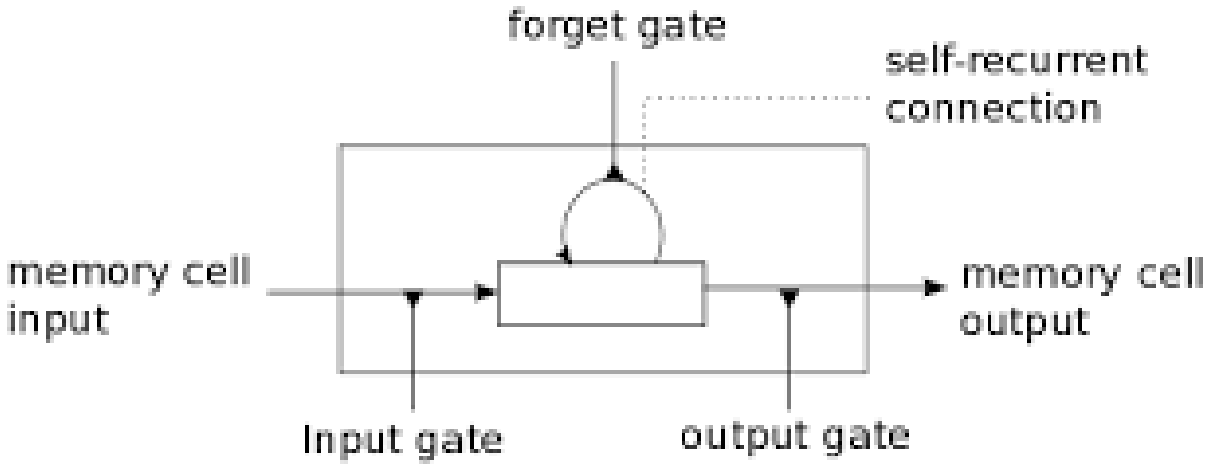
- Compute histogram of tweet timings (binsize = 1 hour)
- Random minute within max hour to tweet
- Bag of Words on timeline tweets
- Select most commonly occurring non-stopword
- We seed the neural network with topics that the user frequently posts about

Leveraging Markov Models



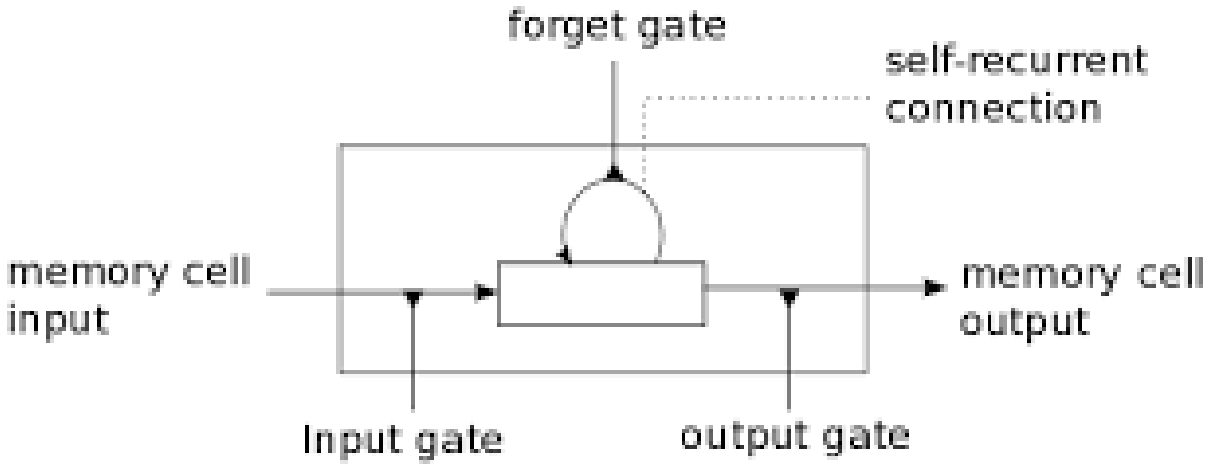
- Popular for text generation: see /r/SubredditSimulator, InfosecTalk TitleBot
- Calculates pairwise frequency of tokens and uses that to generate new ones
- Trained using the most recent posts on the user's timeline

Inside LSTM Neural Networks



LSTM = Long Short Term Memory

Training the Neural Net



LSTM = Long Short Term Memory

- 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

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:

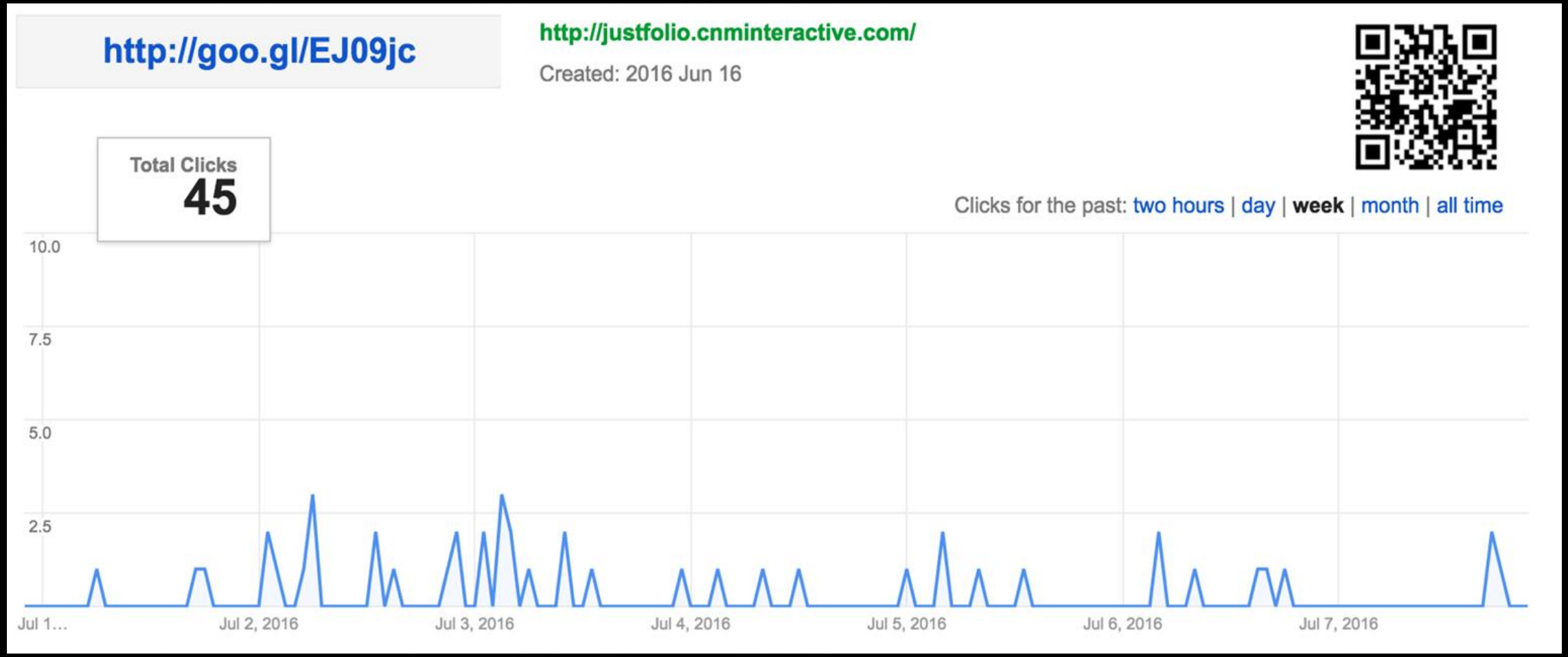
Automated E2E Spear Phishing on Twitter

Here's a malicious URL...

URL:	http://justfolio.cnminteractive.com/
Detection ratio:	6 / 67
Analysis date:	2016-07-06 12:45:13 UTC (7 hours, 48 minutes ago)

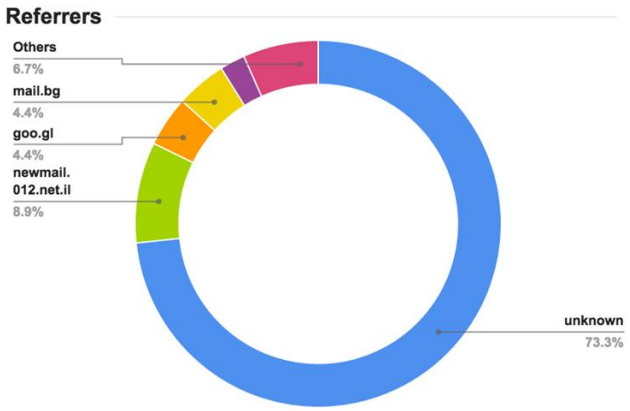
Netcraft	Malicious site
Opera	Malicious site
Sophos	Malicious site
CLEAN MX	Phishing site
Fortinet	Phishing site
Kaspersky	Phishing site

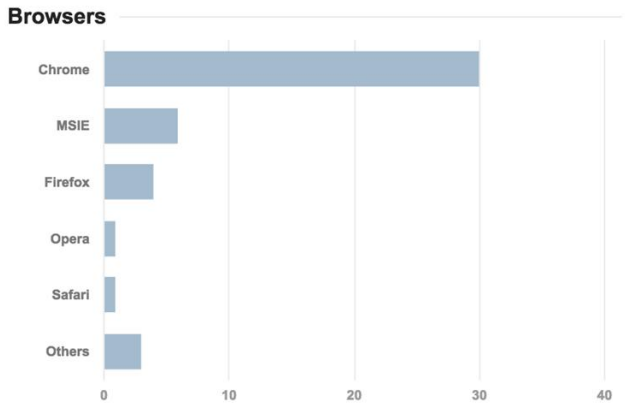
And, apparently goo.gl lets us shorten it!

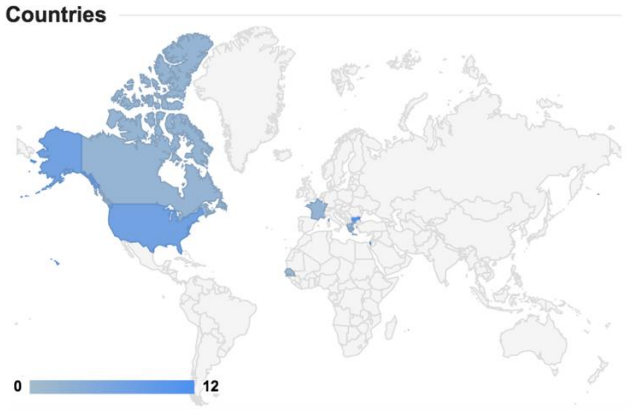


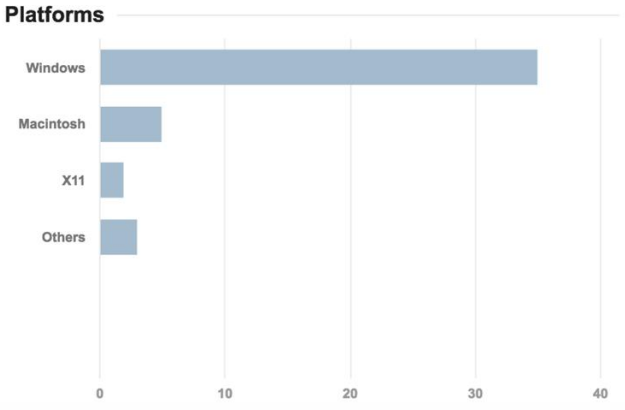
#WeaponizeMe

goo.gl also gives us analytics









#WeaponizeMe

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%

Man vs. Machine

Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

#WeaponizeMe

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

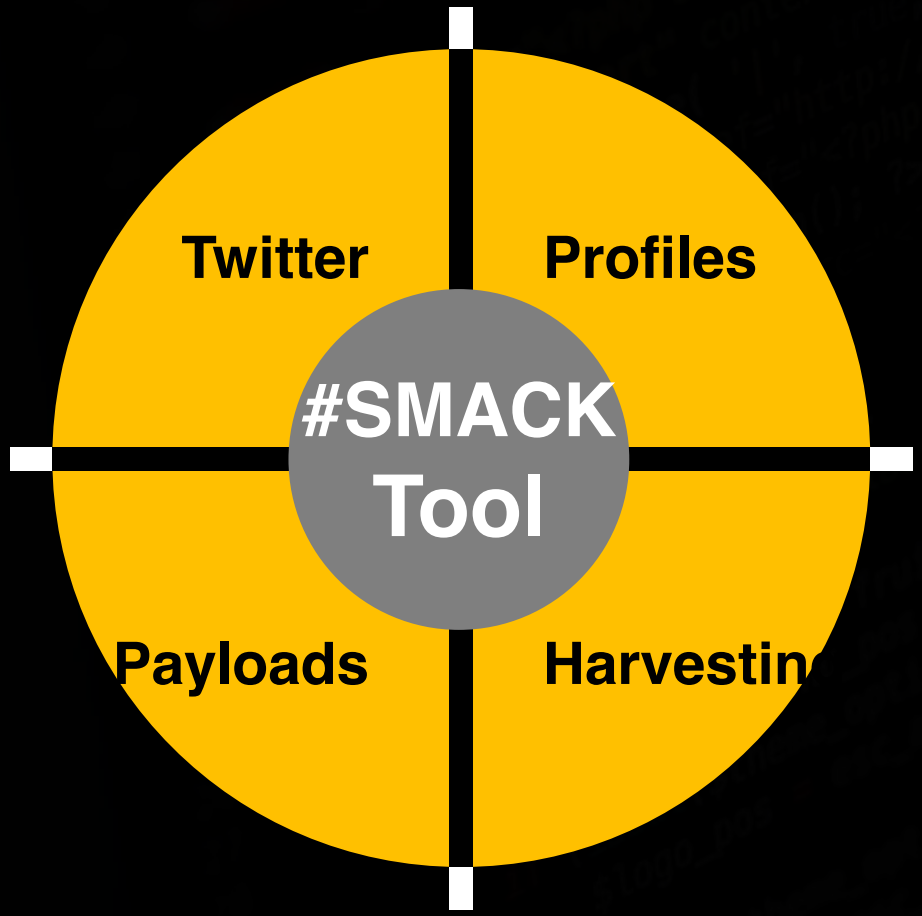
DEMO of Our #SMACK Tool Framework

Wrap Up

Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

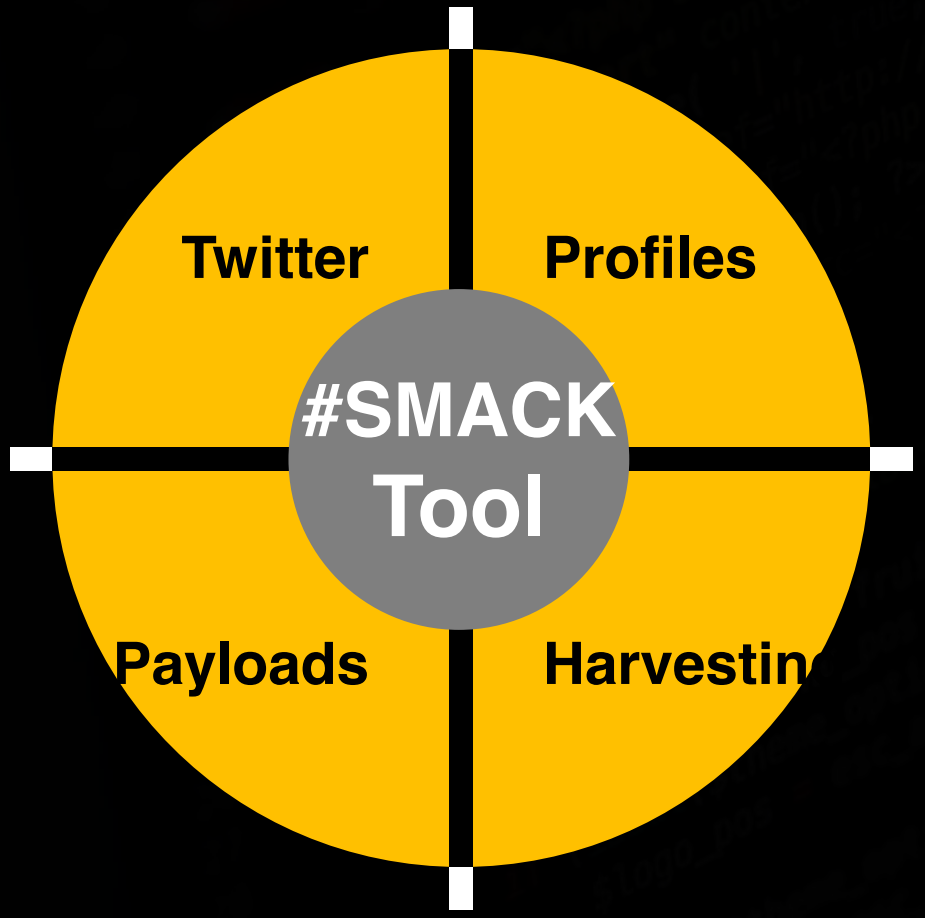
Mitigations



- 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?

Black Hat Sound Bytes



- 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

John Seymour Philip Tully

@_delta_zero @phtully