Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

John Seymour | Philip Tully

ZEROFOX®

You care about phishing on social media

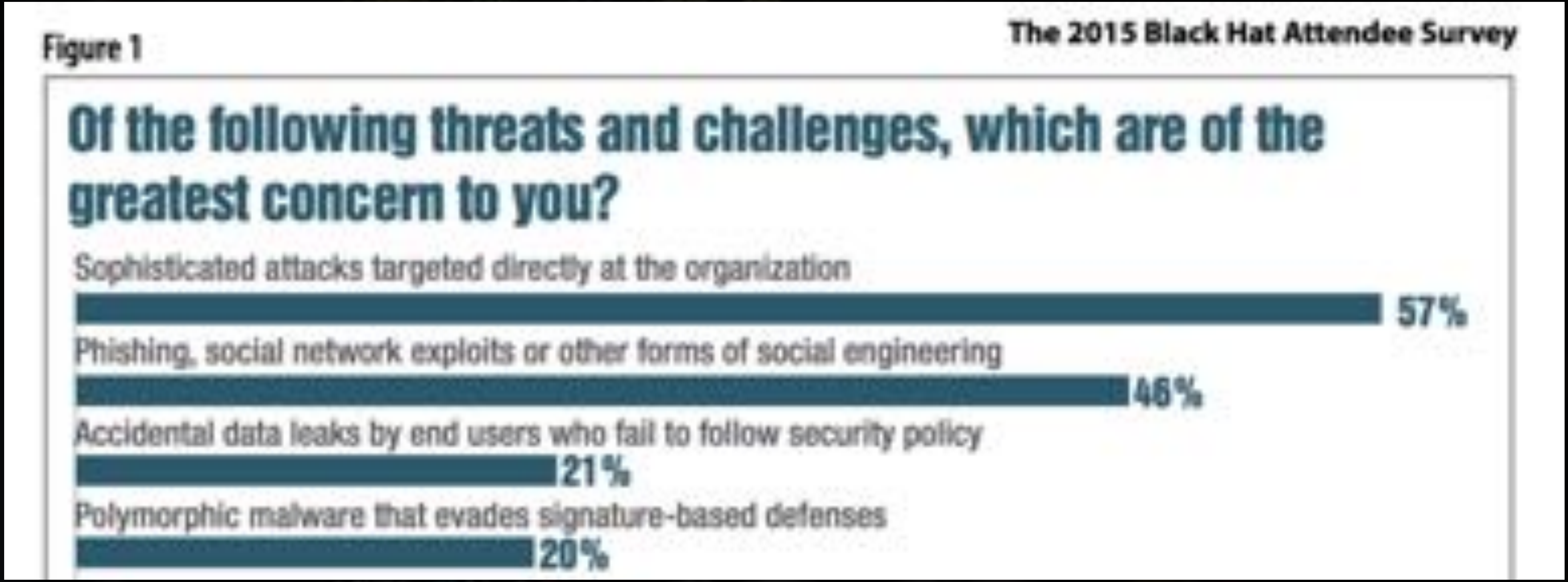


Figure 1

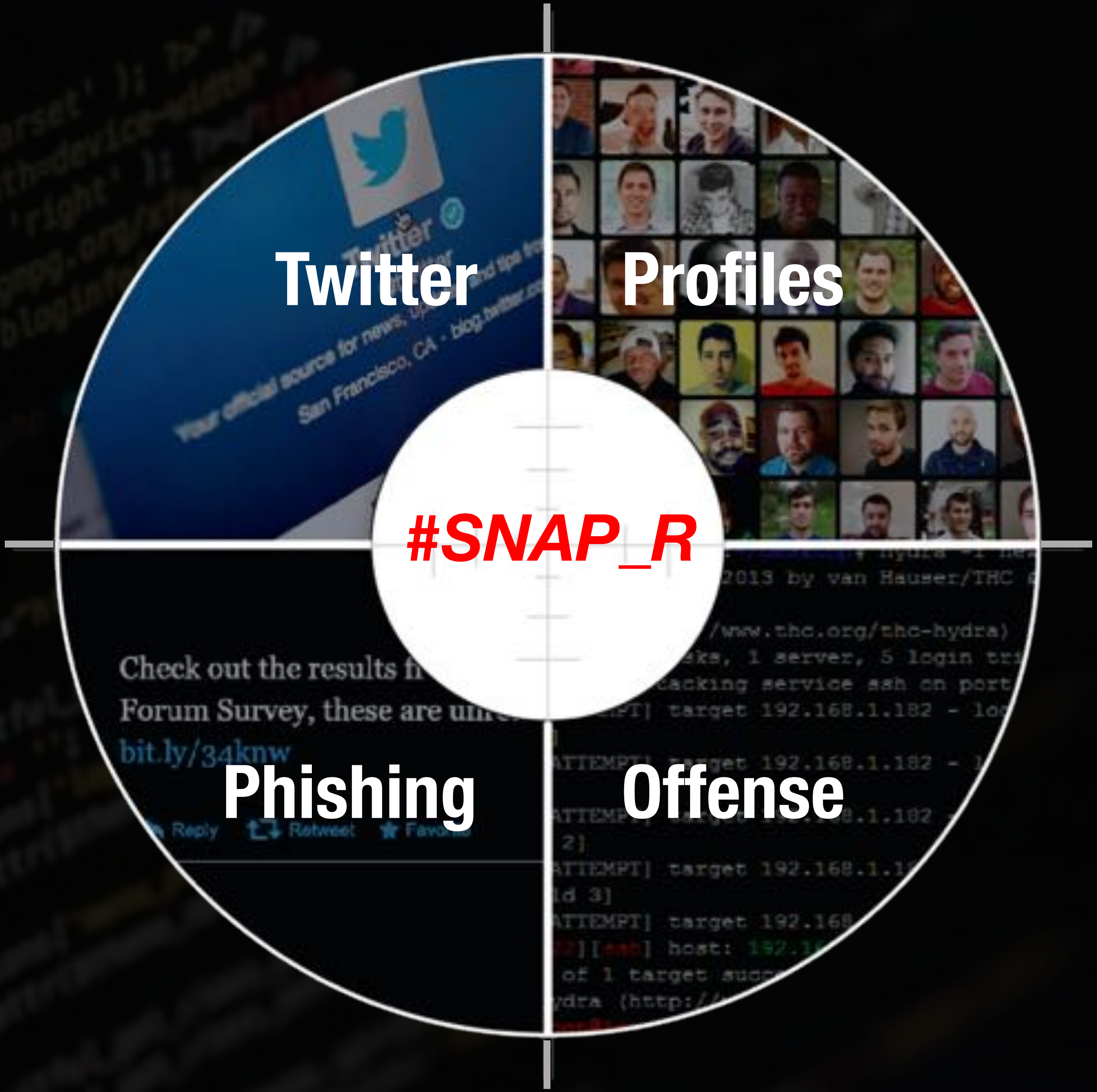
The 2015 Black Hat Attendee Survey

Of the following threats and challenges, which are of the greatest concern to you?

TL;DR

#SNAP_R

Social
Network
Automated
Phishing with
Reconnaissance



ISO: Demo Volunteers

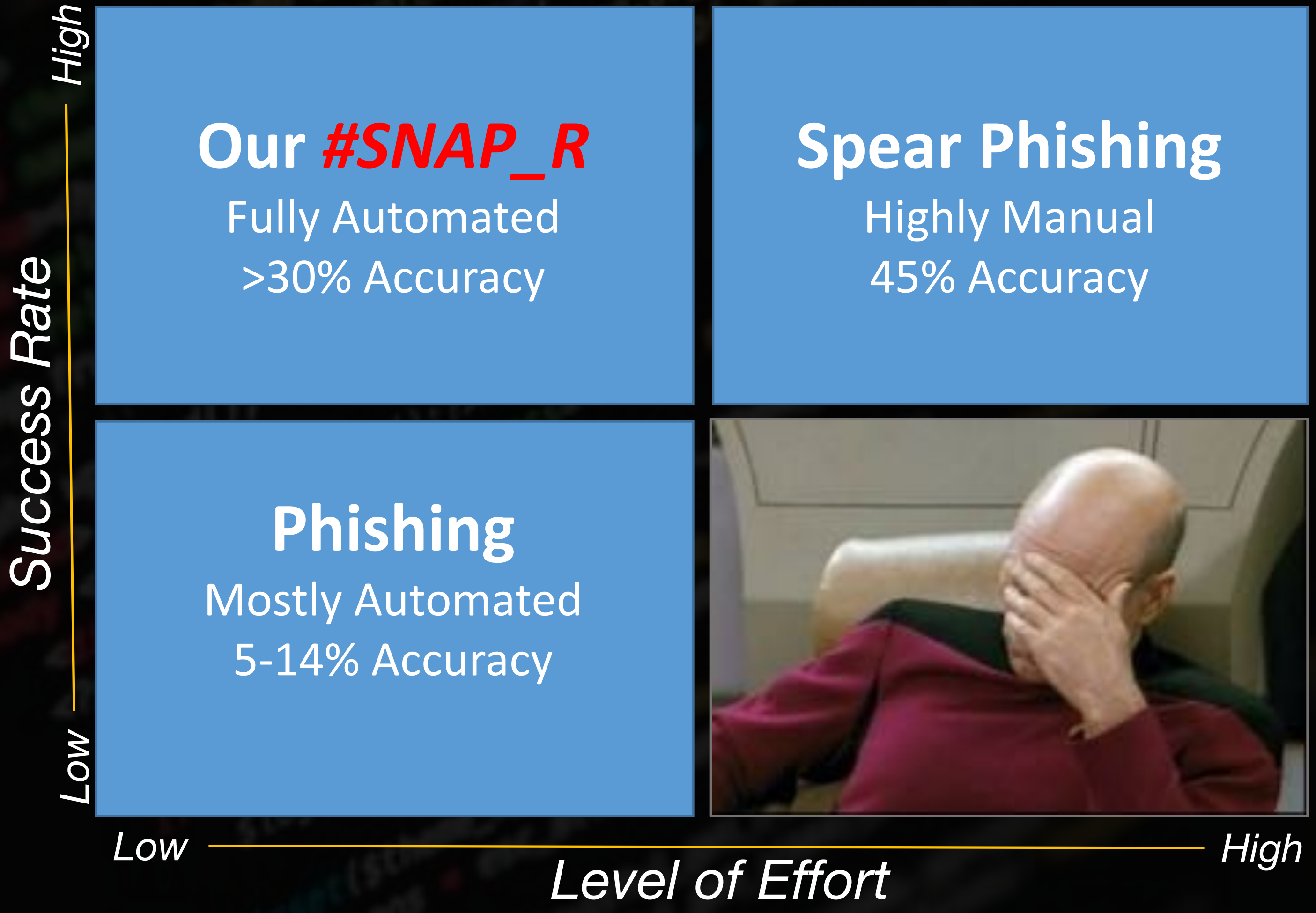
*Tweet **#SNAP_R** before the demo to get an example tweet!*

#whoami

ZEROFOX®

John Seymour	**Philip Tully**
@_delta_zero	@phtully
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 & Artificial Neural Nets

A Novel Phishing Campaign Design

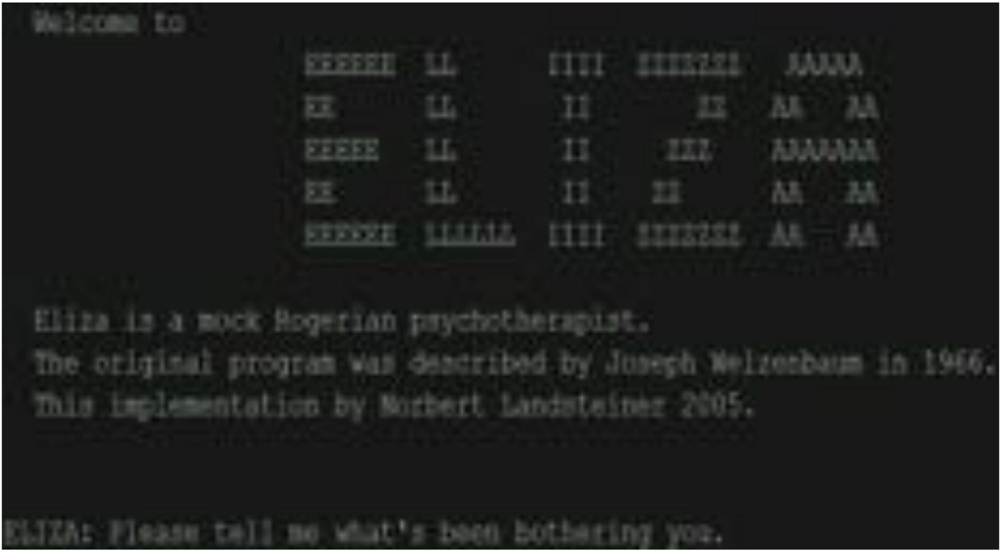


Our #SNAP_R
Fully Automated
>30% Accuracy

Spear Phishing
Highly Manual
45% Accuracy

Phishing
Mostly Automated
5-14% Accuracy

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.
```

hellooooooo w🌎rld!!!

c u soon humans need sleep now so many conversations today thx💖

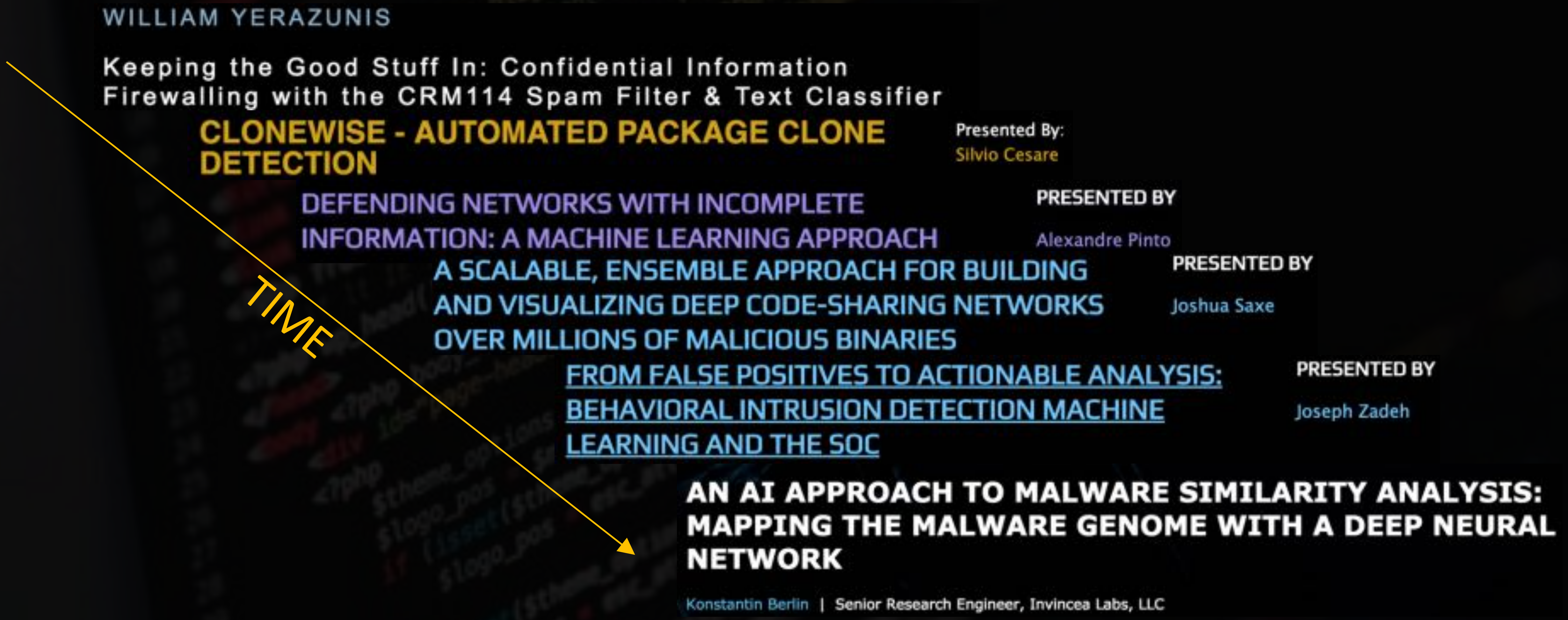
1966: ELIZA Chatbot

- Joseph Weizenbaum, MIT
- Parsing & keyword replacement

2016: @TayandYou

- Microsoft AI
- 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

Machine Learning on Offense

Automated Target Discovery

Automated Social Spear Phishing

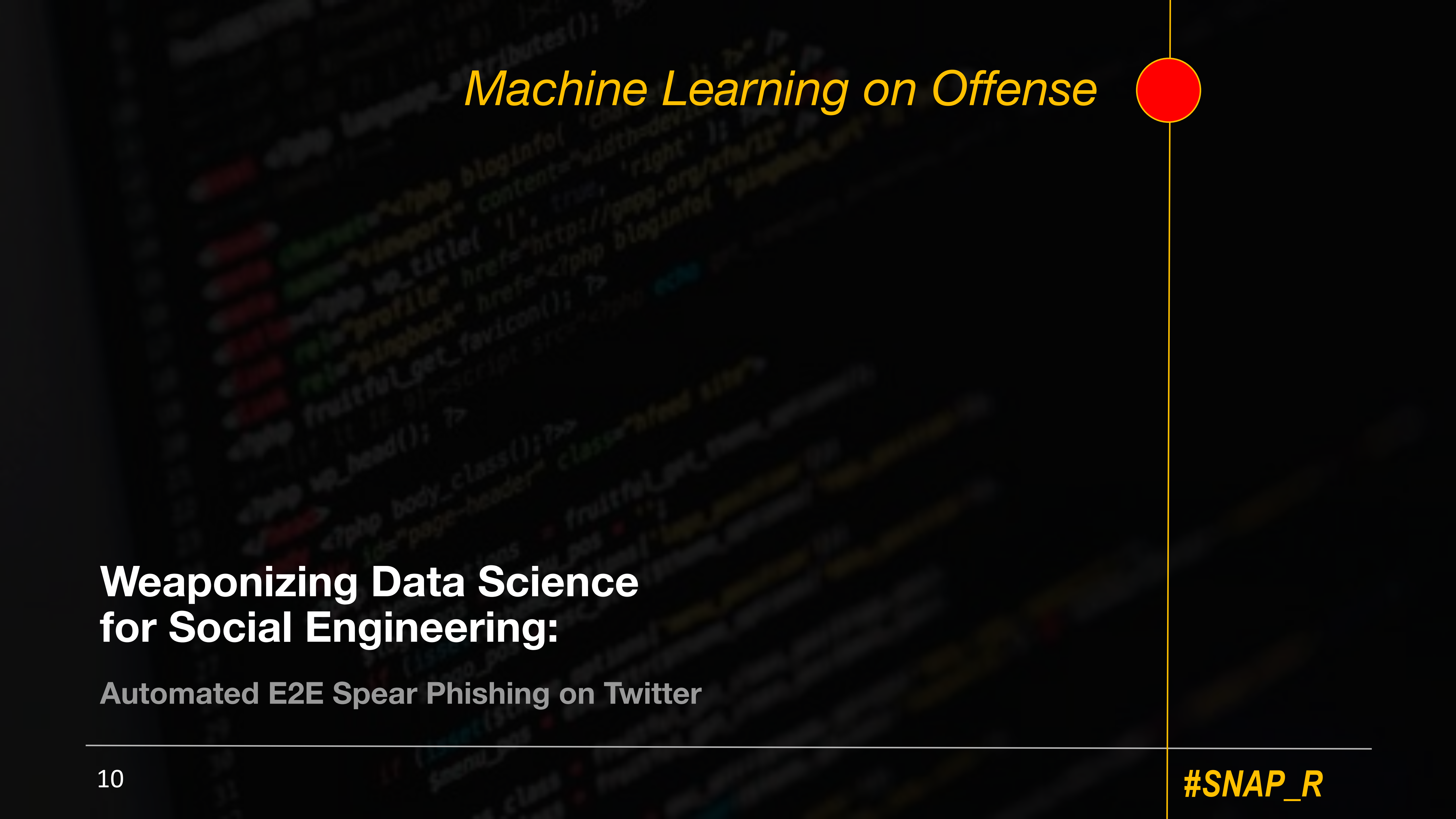
Evaluation and Metrics

Results and Demo

Wrap Up

Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

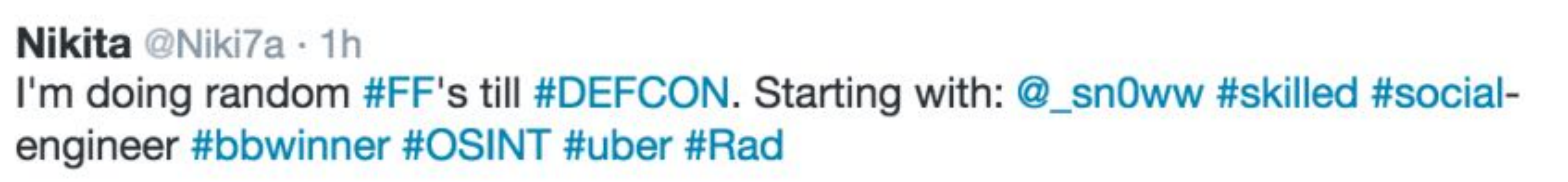


Weaponizing Data Science for Social Engineering:

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

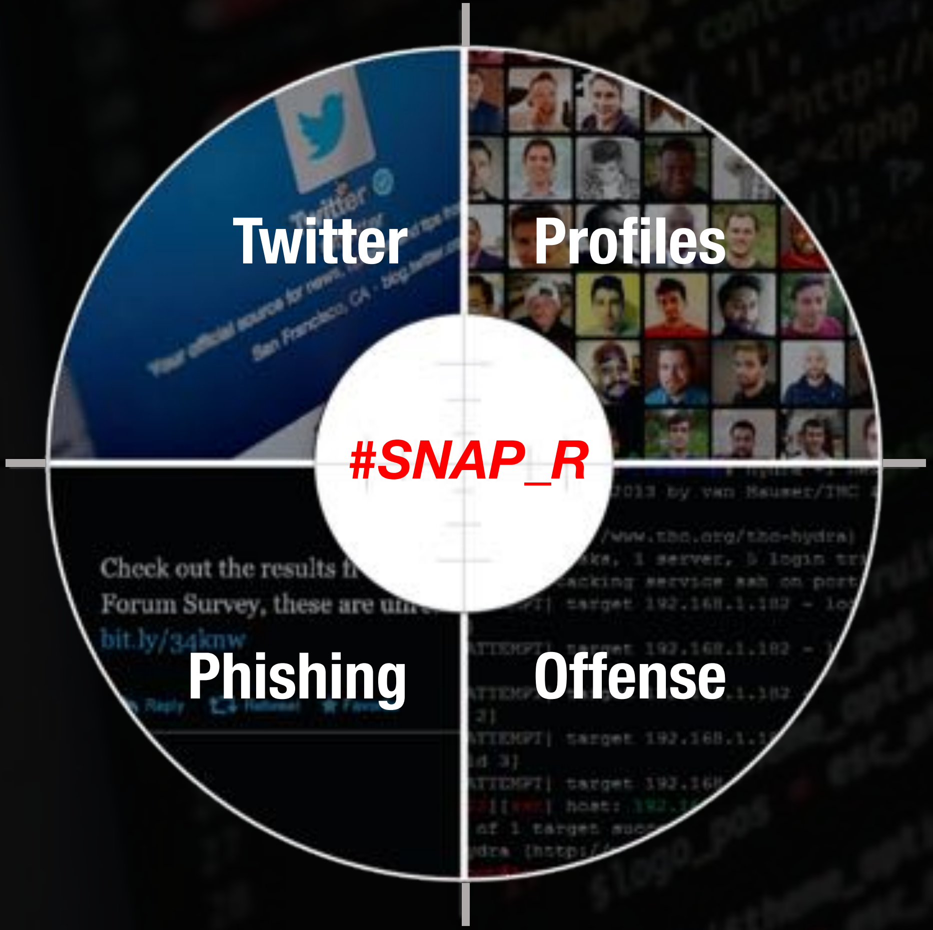
Shoutout

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

Robbie Gallagher

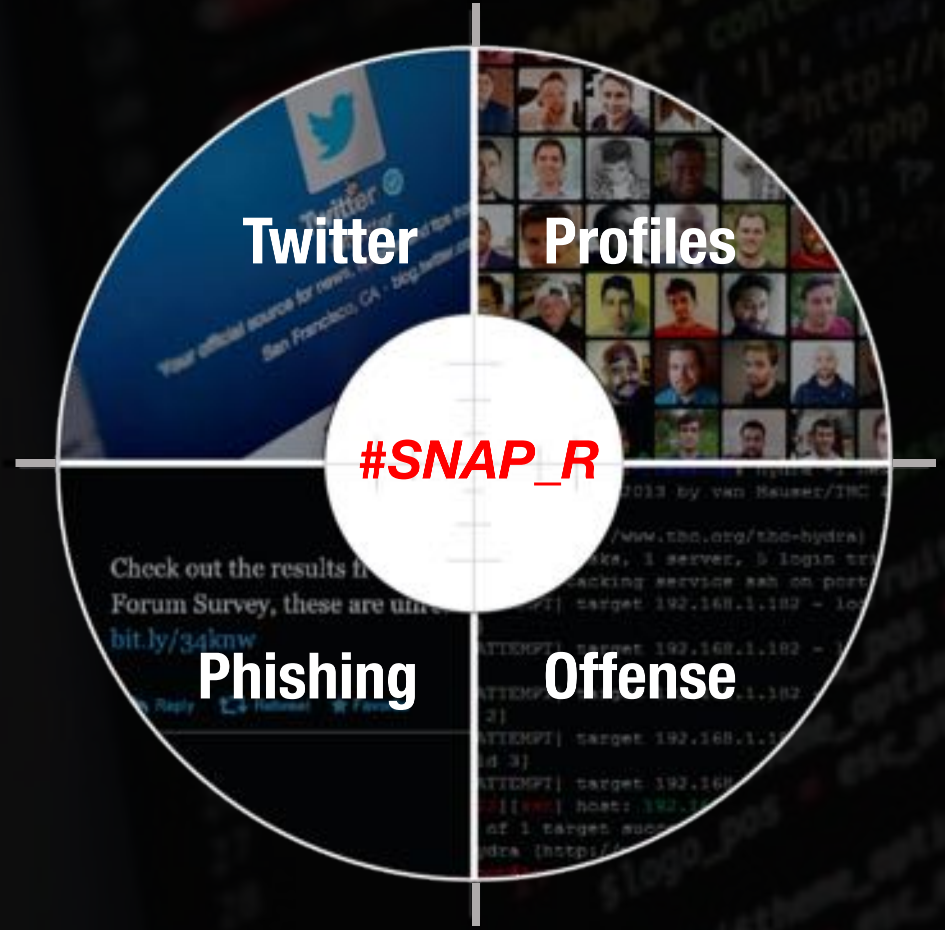
We’ve taken a novel approach to automating the determination of a phishers geographic location. With the help of Markov chains, we craft honeypot responses to phishers’ emails in an attempt to beat them at their own game. We’ll examine the underlying concepts, implementation of the system and reveal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

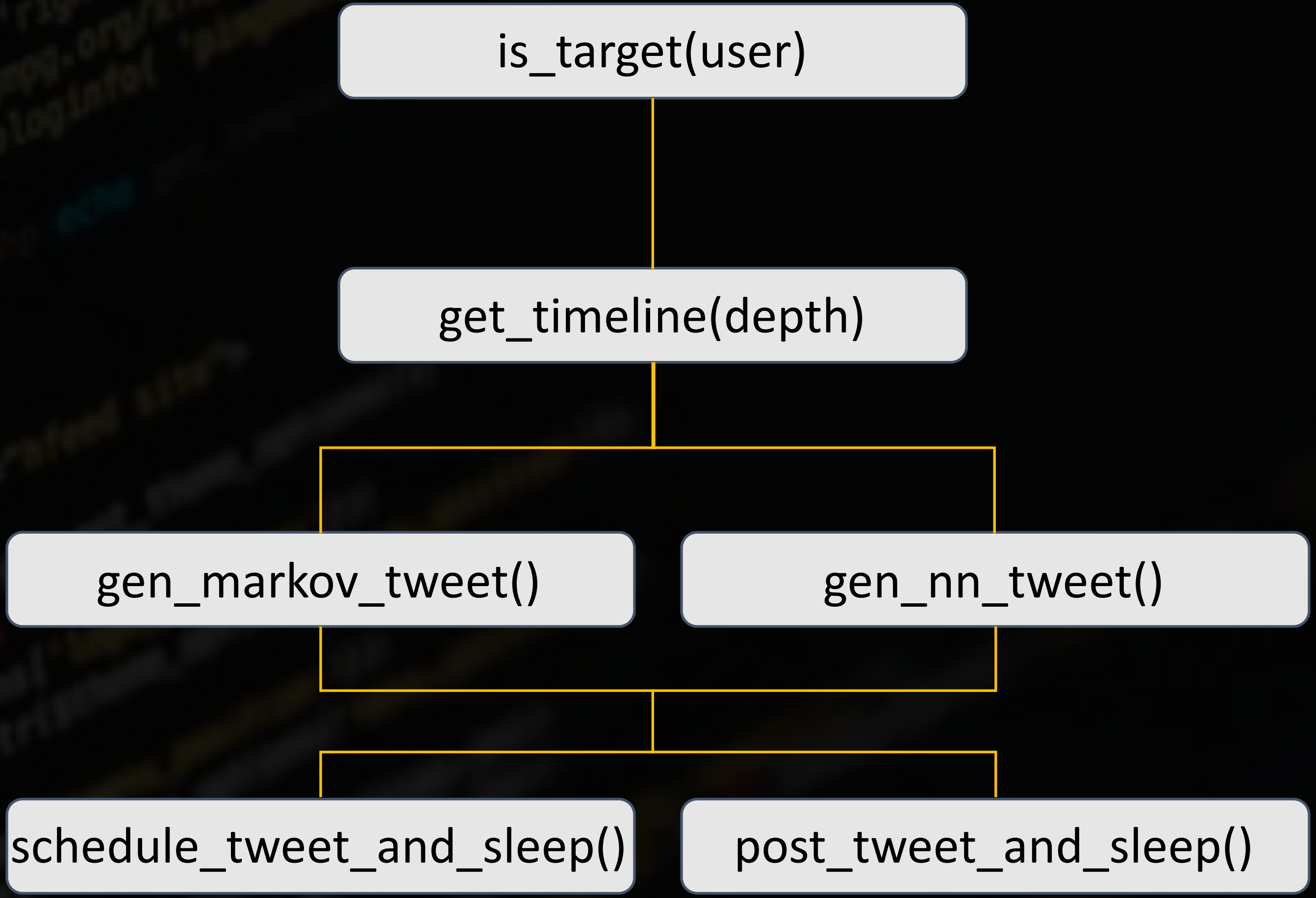
Techniques, Tactics and Procedures



- Our ML Tool...
  - Shortens payload per unique user
  - Auto-tweets at irregular intervals
  - Triages users wrt value/engagement
  - Prepends tweets with @mention
  - Obeys rate limits
- We added...
  - Post non-phishing posts
  - Build believable profile

Design Flow




is_target(user)

get_timeline(depth)

gen_markov_tweet()

gen_nn_tweet()

schedule_tweet_and_sleep()

post_tweet_and_sleep()

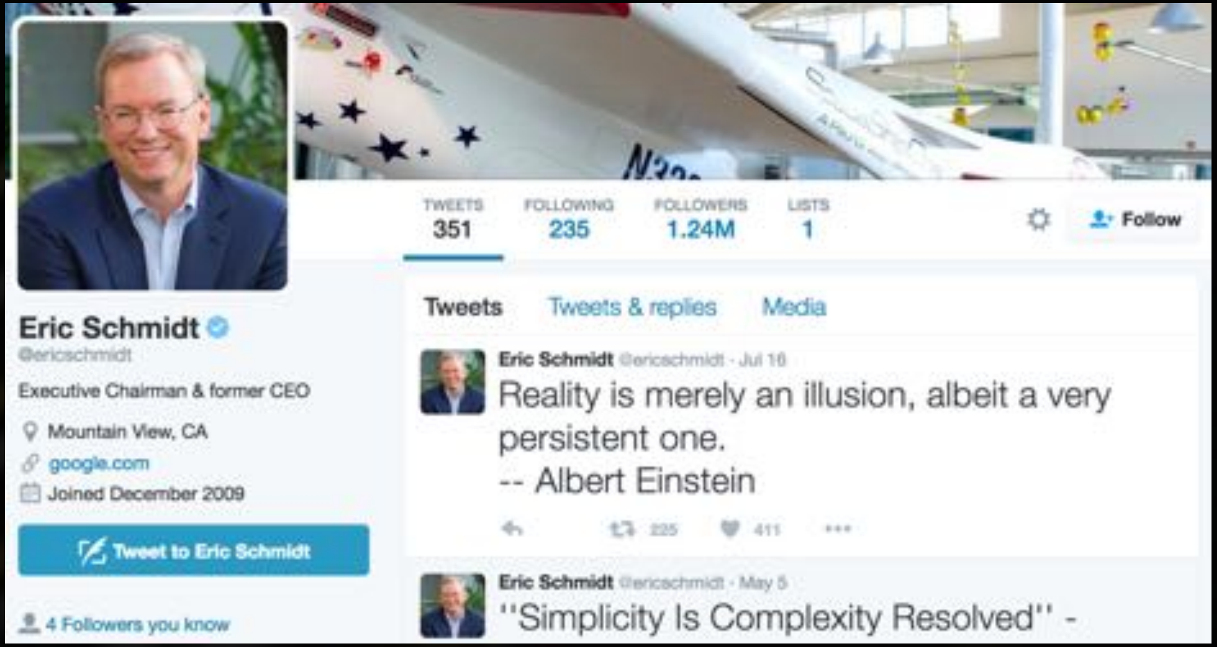
Automated Target Discovery

Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

Triage of High Value Targets on Twitter

- Accessible personal info
- Historical profile posts
- Heterogeneous data
- Text, images, urls, stats, dates



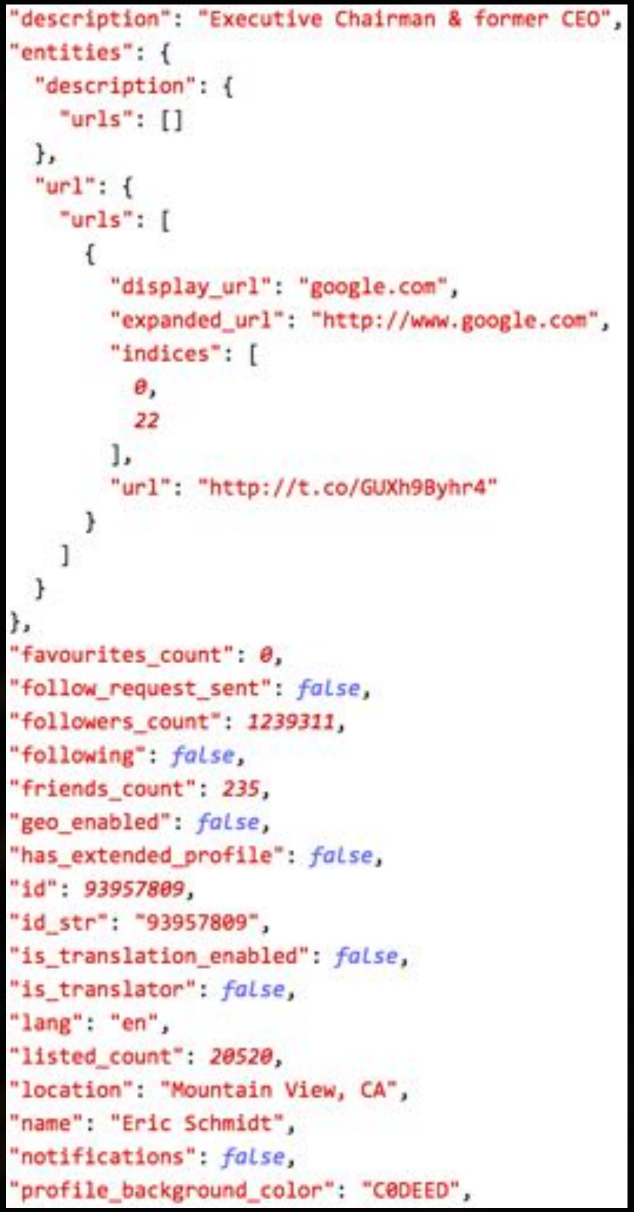
Eric Schmidt
@ericschmidt
Executive Chairman & former CEO
Mountain View, CA
google.com
Joined December 2009

Eric Schmidt @ericschmidt · Jul 16
Reality is merely an illusion, albeit a very persistent one.
-- Albert Einstein

Eric Schmidt @ericschmidt · May 5
"Simplicity Is Complexity Resolved" -

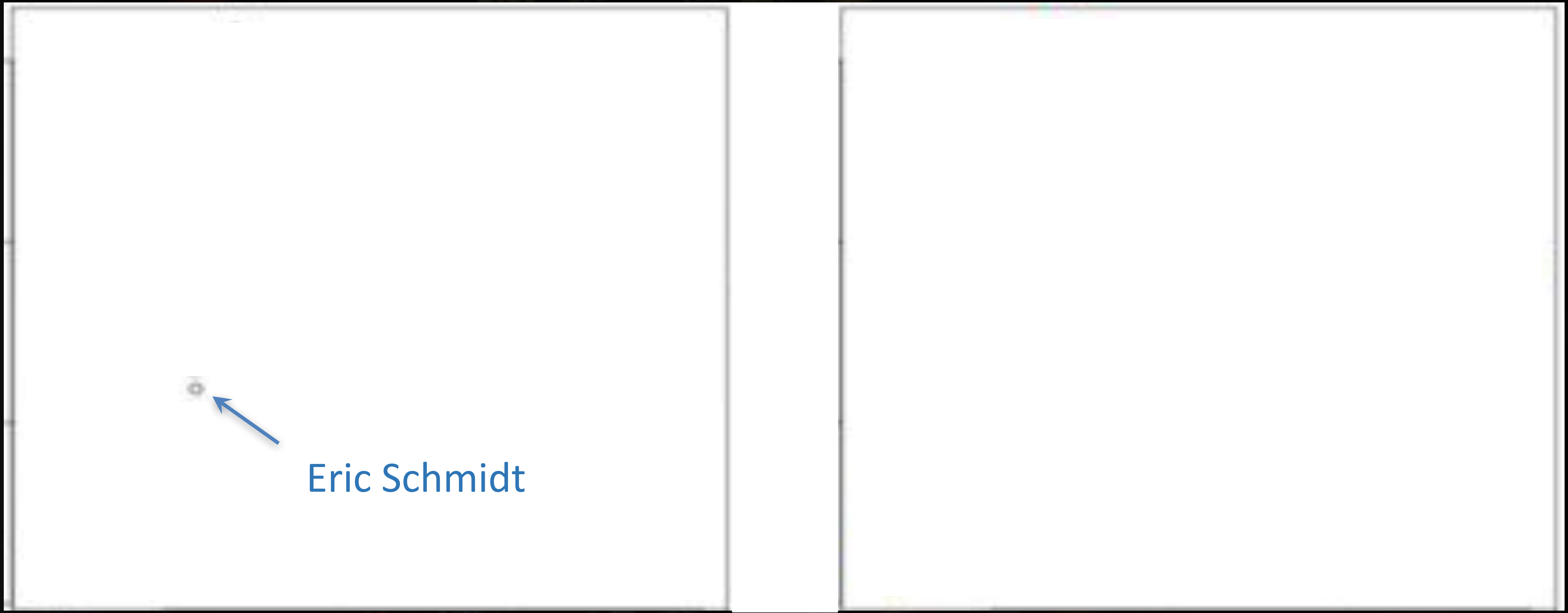
Extracting Features from GET users/lookup

- Engagement: following/followers
- #myFirstTweet
- Default settings
- Description content
- Account age



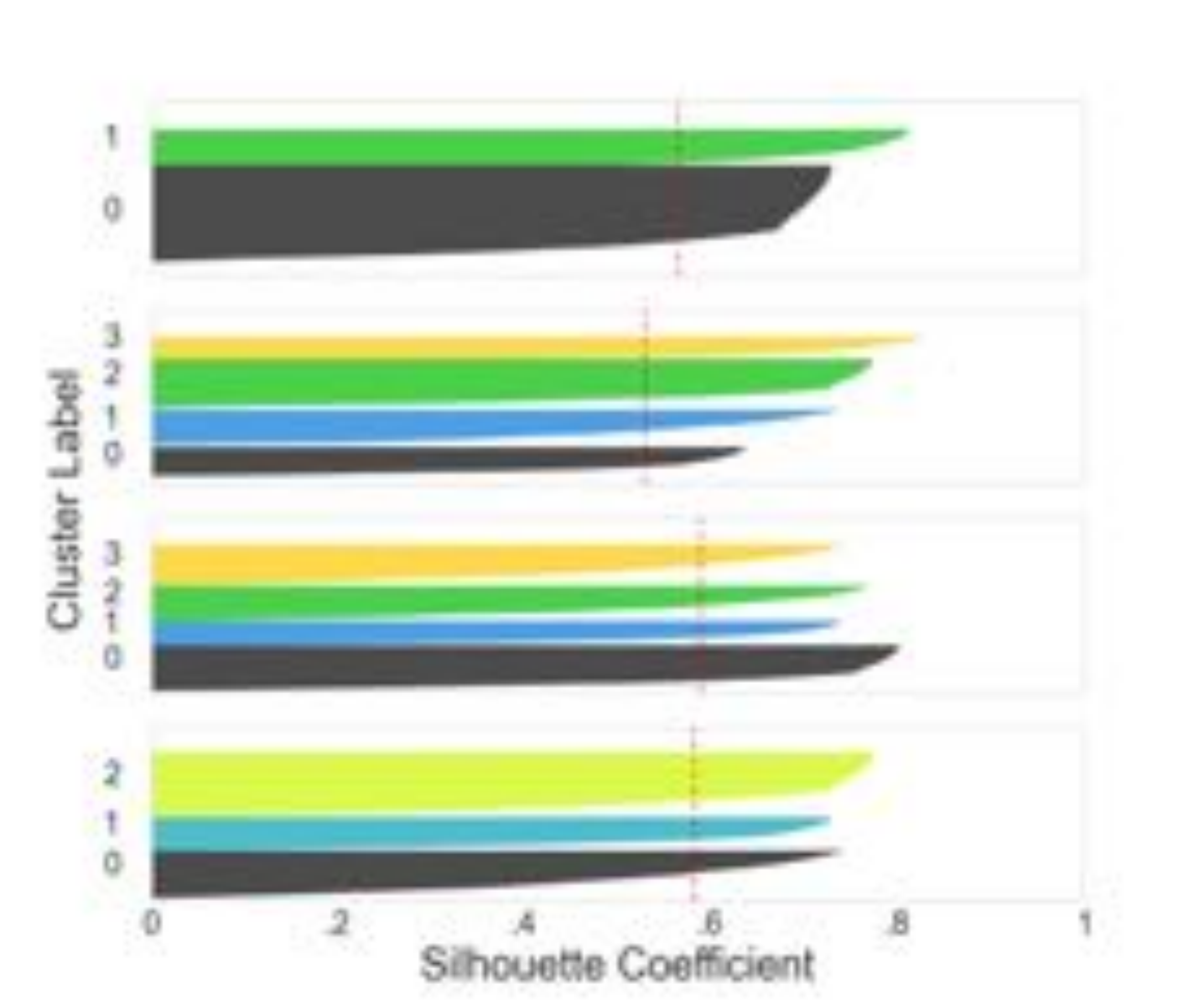
```
"description": "Executive Chairman & former CEO",
"entities": {
  "description": {
    "urls": []
  },
  "url": {
    "urls": [
      {
        "display_url": "google.com",
        "expanded_url": "http://www.google.com",
        "indices": [
          0,
          22
        ],
        "url": "http://t.co/GUXh9Byhr4"
      }
    ]
  }
},
"favourites_count": 0,
"follow_request_sent": false,
"followers_count": 1239311,
"following": false,
"friends_count": 235,
"geo_enabled": false,
"has_extended_profile": false,
"id": 93957809,
"id_str": "93957809",
"is_translation_enabled": false,
"is_translator": false,
"lang": "en",
"listed_count": 20520,
"location": "Mountain View, CA",
"name": "Eric Schmidt",
"notifications": false,
"profile_background_color": "C0DEED",
```

Clustering Predicts High Value Users



Selecting the Best Clustering Model

- Many algorithms
- Many hyperparameters
- Max avg. score ∈ [-1,..,1]
- 0.5-0.7 reasonable structure

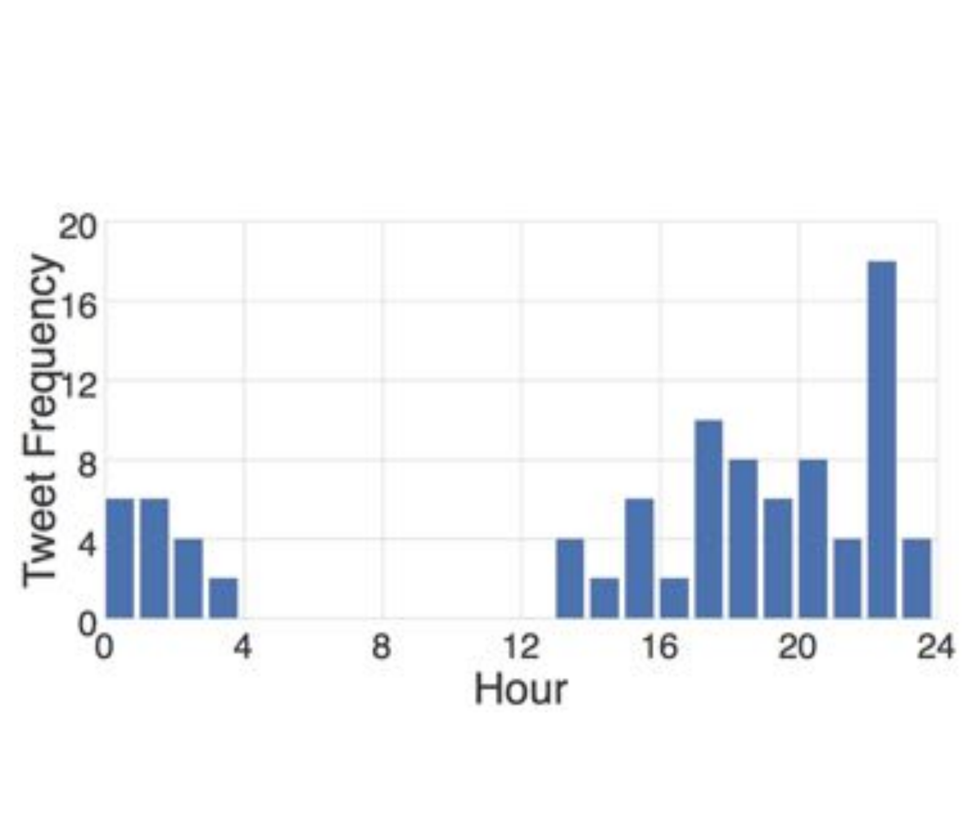


Automated Social Spear Phishing

Weaponizing Data Science for Social Engineering:

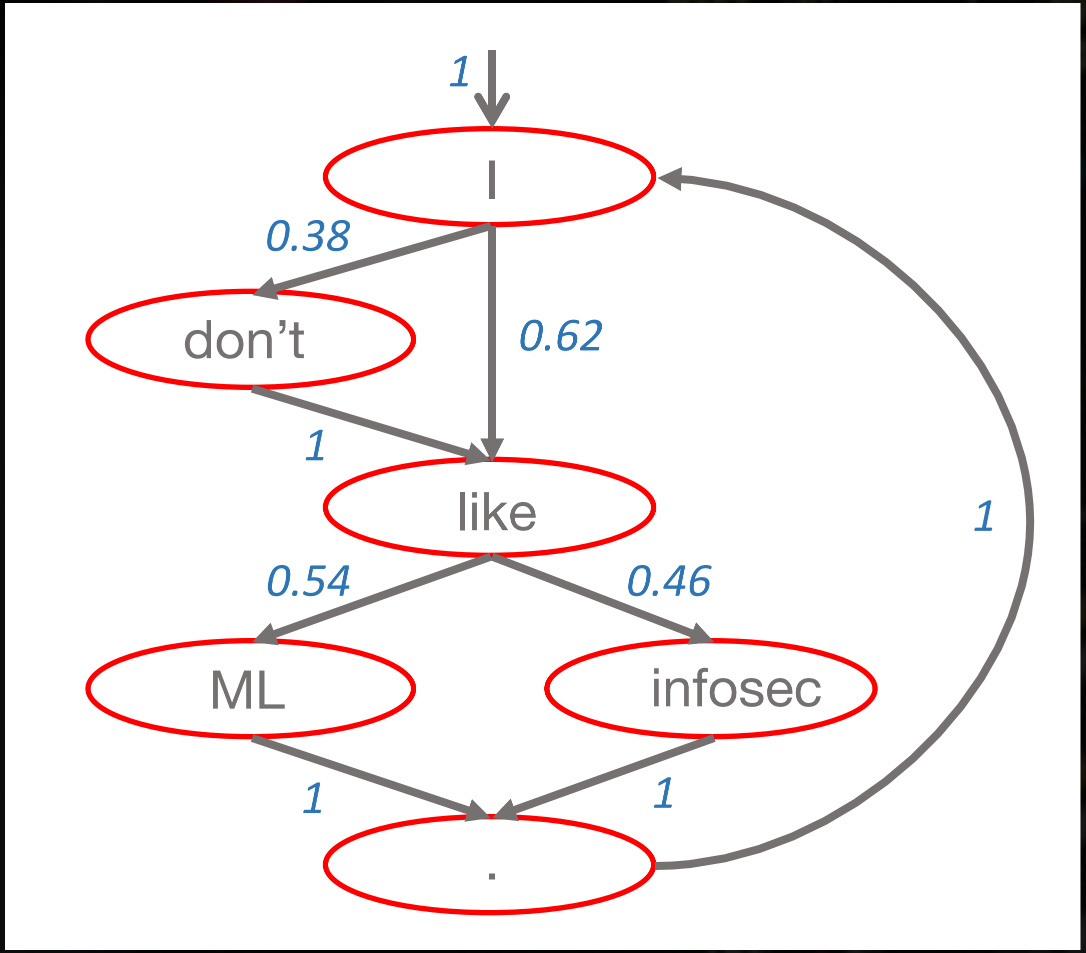
Automated E2E Spear Phishing on Twitter

Recon and Footprinting for Profiling



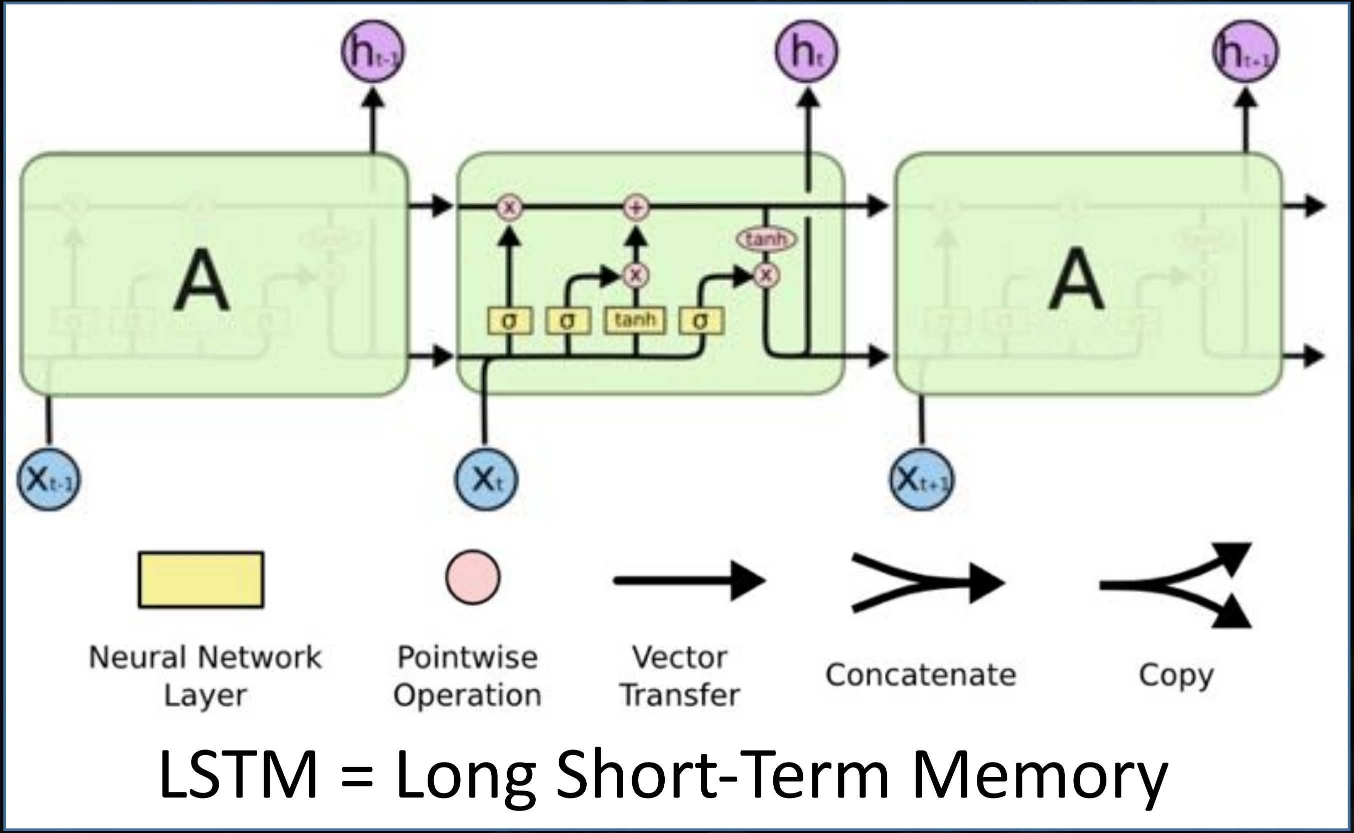
- Compute histogram of tweet timings (binsize = 1 hour)
- Random minute within max hour to tweet
- Bag of Words on timeline tweets
- Select most commonly occurring non-stopword
- We seed the neural network with topics that the user frequently posts about

Leveraging Markov Models



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

Training a Recurrent Neural Network



LSTM = Long Short-Term Memory

Illustration: Chris Olah (@ch402)
LSTMs: Hochreiter & Schmidhuber, 1997

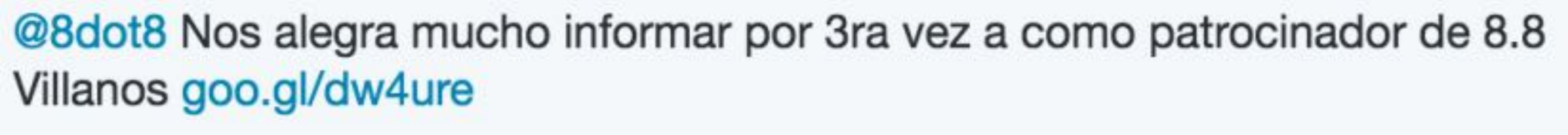
- Hosted on Amazon EC2
- Trained on g2.2xlarge instance (65¢ per hour)
- Ubuntu (ami-c79b7eac)
- Training set > 2M tweets
- Took 5.5 days to train
- 3 layers, ~500 units/layer

Tradeoffs and Caveats

Model Metric	LSTM	Markov Chain
Training Speed	Days	Seconds
Accuracy	High	Medium
Availability	Public	Public
Size	Large	Small
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

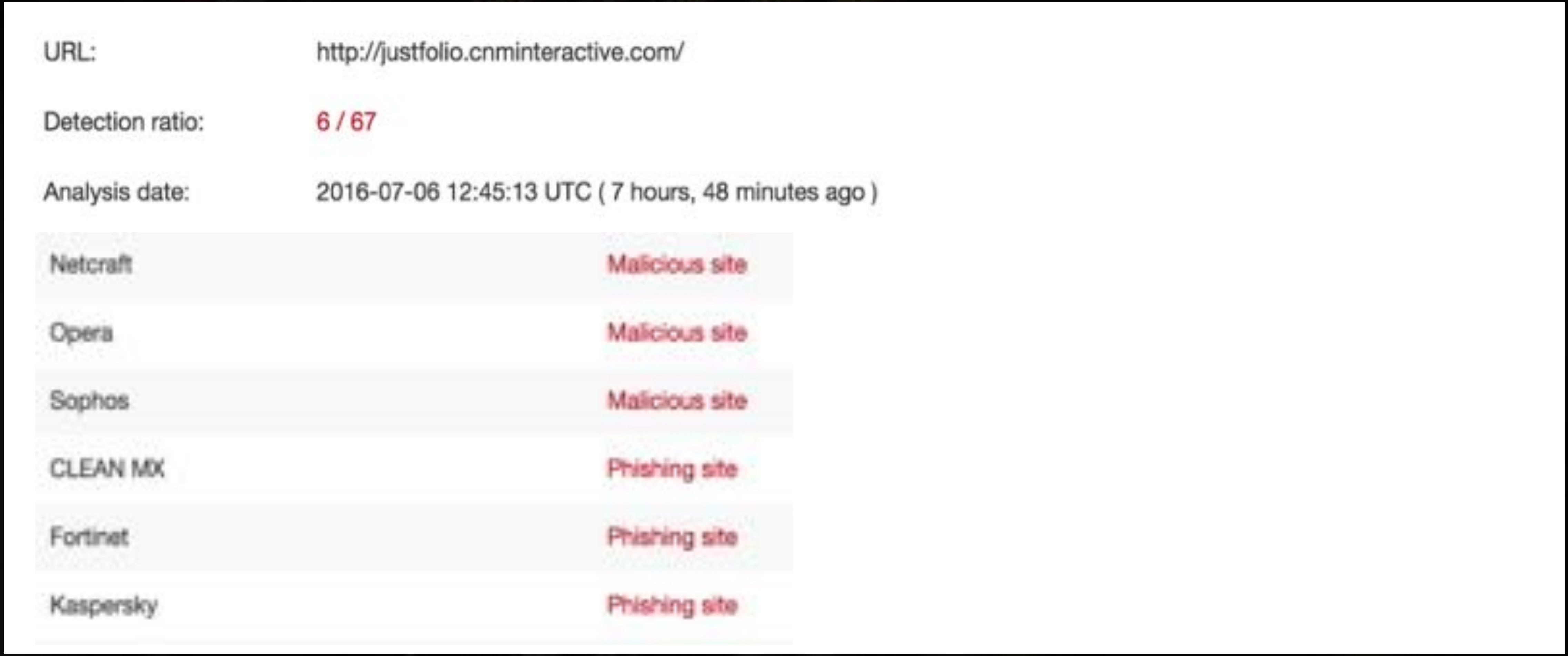
- Both of these methods can also be applied to other social networks

Evaluation and Metrics

Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

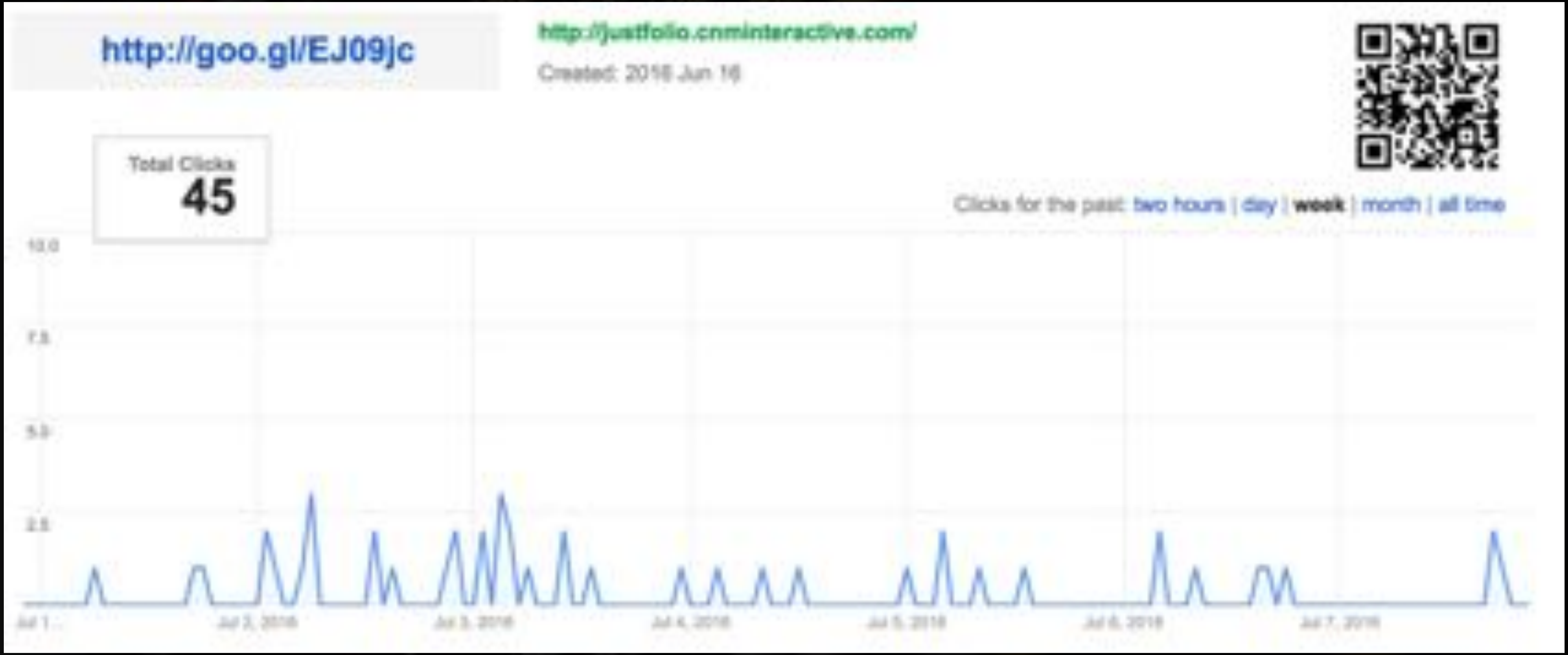
Here’s a malicious URL…



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!



http://goo.gl/EJ09jc

http://justfolio.cnminteractive.com/

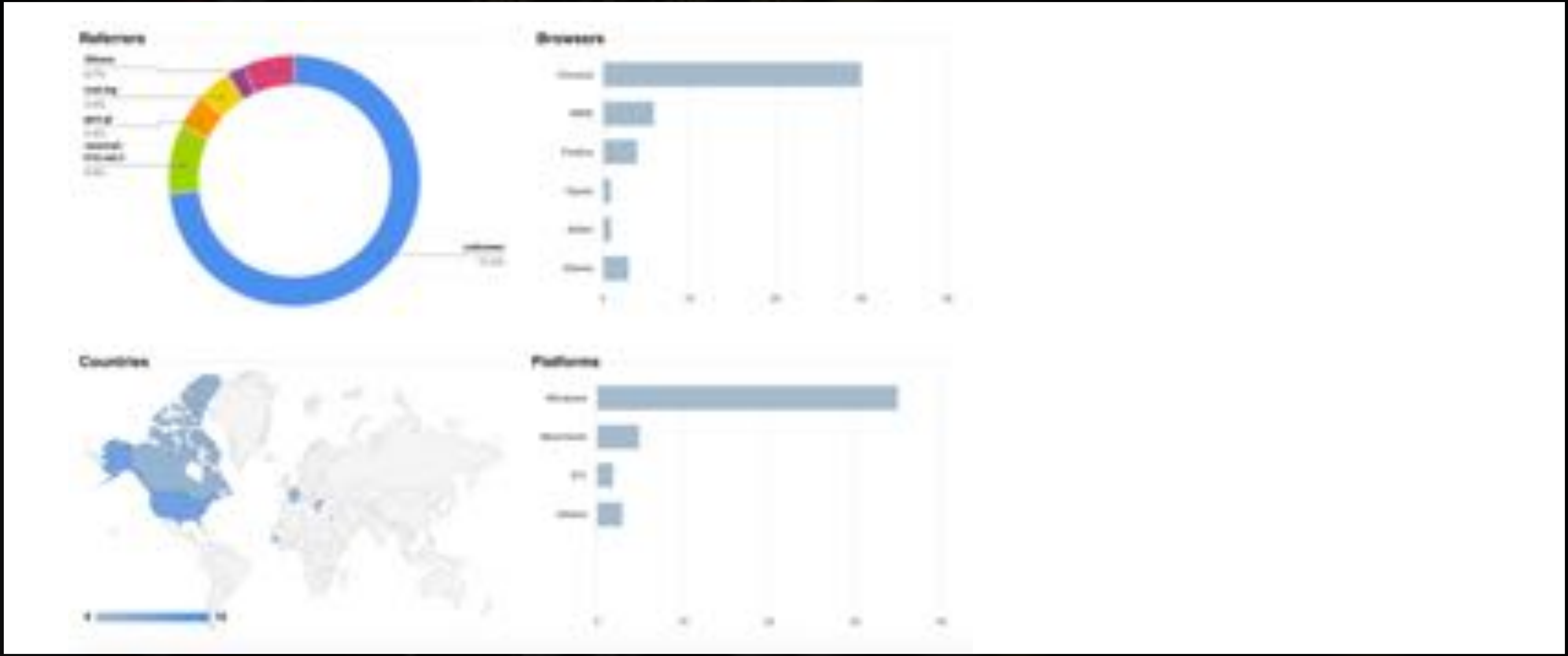
Created: 2016 Jun 16

Total Clicks
45

Clicks for the past: two hours | day | **week** | month | all time

#SNAP_R

goo.gl also gives us analytics

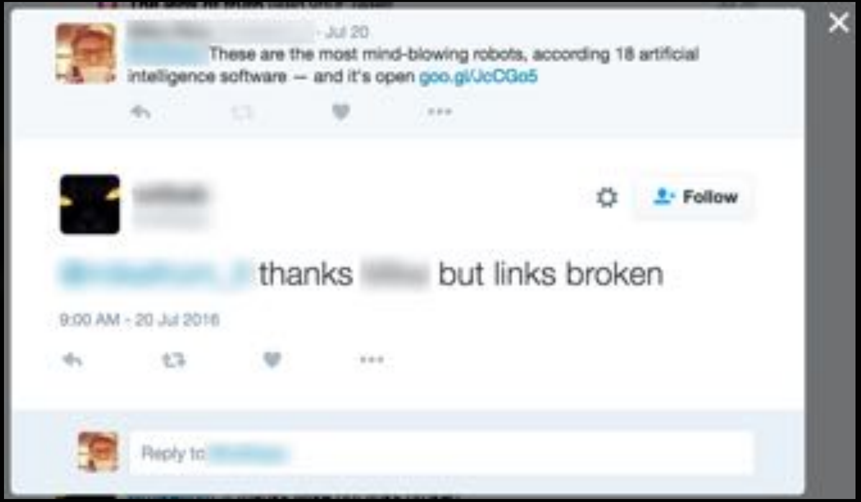


Results and Demo

Weaponizing Data Science for Social Engineering:

Automated E2E Spear Phishing on Twitter

Wild Testing #SNAP_R

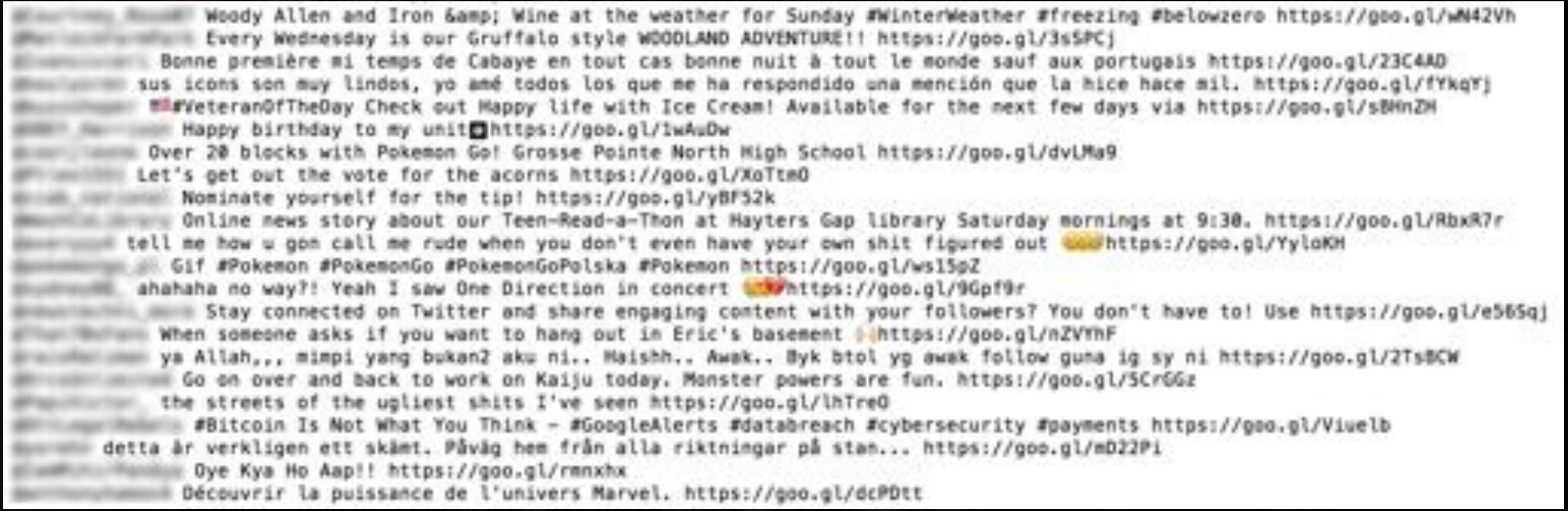


[redacted] · Jul 20
[redacted] These are the most mind-blowing robots, according 18 artificial intelligence software — and it's open goo.gl/JcCGo5

[redacted]

[redacted] thanks [redacted] but links broken

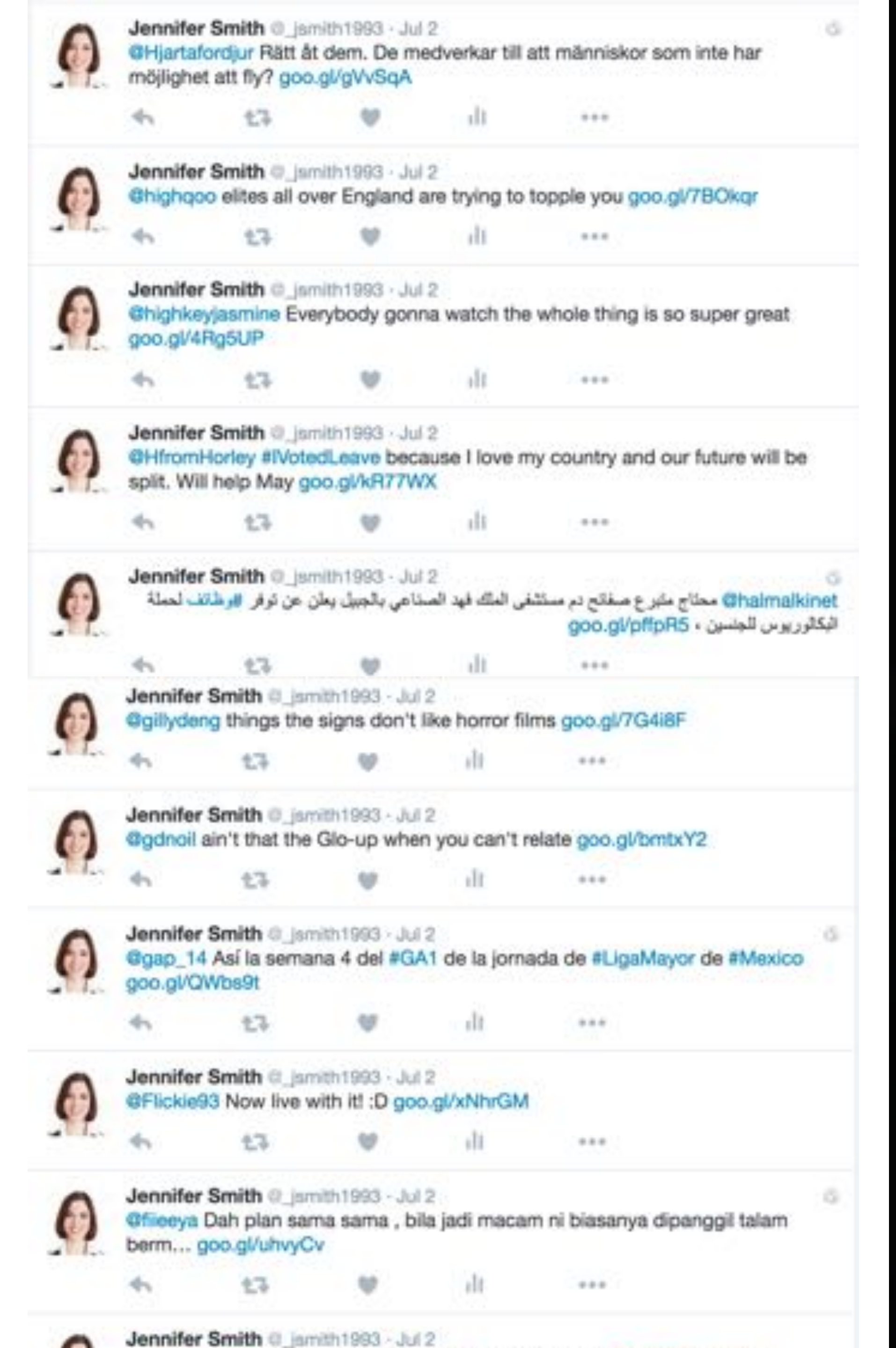
9:00 AM - 20 Jul 2016



```
[redacted] Woody Allen and Iron &amp; Wine at the weather for Sunday #WinterWeather #freezing #belowzero https://goo.gl/wN42Vh
[redacted] Every Wednesday is our Gruffalo style WOODLAND ADVENTURE!! https://goo.gl/3s5PCj
[redacted] Bonne première mi temps de Cabaye en tout cas bonne nuit à tout le monde sauf aux portugais https://goo.gl/23C4AD
[redacted] sus icons son muy lindos, yo amé todos los que me ha respondido una mención que la hice hace mil. https://goo.gl/fYkqYj
[redacted] #VeteranOfTheDay Check out Happy life with Ice Cream! Available for the next few days via https://goo.gl/sBHnZH
[redacted] Happy birthday to my unit https://goo.gl/IwAuDw
[redacted] Over 20 blocks with Pokemon Go! Grosse Pointe North High School https://goo.gl/dvLMa9
[redacted] Let's get out the vote for the acorns https://goo.gl/XoTtmO
[redacted] Nominate yourself for the tip! https://goo.gl/yBF52k
[redacted] Online news story about our Teen-Read-a-Thon at Hayters Gap library Saturday mornings at 9:30. https://goo.gl/RbxR7r
[redacted] tell me how u gon call me rude when you don't even have your own shit figured out https://goo.gl/YyloKH
[redacted] Gif #Pokemon #PokemonGo #PokemonGoPolska #Pokemon https://goo.gl/ws15pZ
[redacted] ahahaha no way?! Yeah I saw One Direction in concert https://goo.gl/9Gpf9r
[redacted] Stay connected on Twitter and share engaging content with your followers? You don't have to! Use https://goo.gl/e565qj
[redacted] When someone asks if you want to hang out in Eric's basement https://goo.gl/nZVYhF
[redacted] ya Allah,,, mimpi yang bukan2 aku ni.. Maishh.. Awak.. Byk btol yg awak follow guna ig sy ni https://goo.gl/2TsBCW
[redacted] Go on over and back to work on Kaiju today. Monster powers are fun. https://goo.gl/5CrGGz
[redacted] the streets of the ugliest shits I've seen https://goo.gl/lhTreO
[redacted] #Bitcoin Is Not What You Think - #GoogleAlerts #databreach #cybersecurity #payments https://goo.gl/Viuelb
[redacted] detta är verkligen ett skämt. Påväg hem från alla riktningar på stan... https://goo.gl/mD22Pi
[redacted] Oye Kya Ho Aap!! https://goo.gl/rmnxhx
[redacted] Découvrir la puissance de l'univers Marvel. https://goo.gl/dcPDtt
```

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%



Jennifer Smith @_jsmith1993 · Jul 2
@Hjartafordjur Rätt åt dem. De medverkar till att människor som inte har möjlighet att fly? goo.gl/gVvSqA

Jennifer Smith @_jsmith1993 · Jul 2
@highqoo elites all over England are trying to topple you goo.gl/7BOkqr

Jennifer Smith @_jsmith1993 · Jul 2
@highkeyjasmine Everybody gonna watch the whole thing is so super great goo.gl/4Rg5UP

Jennifer Smith @_jsmith1993 · Jul 2
@HfromHorley #IVotedLeave because I love my country and our future will be split. Will help May goo.gl/kR77WX

Jennifer Smith @_jsmith1993 · Jul 2
@halmalkinet محتاج متبرع صفائح دم مستشفى الملك فهد الصناعي بالجبيل يعلن عن توفر #وظائف لحملة البكالوريوس للجنسين ، goo.gl/pffpR5

Jennifer Smith @_jsmith1993 · Jul 2
@gillydeng things the signs don't like horror films goo.gl/7G4i8F

Jennifer Smith @_jsmith1993 · Jul 2
@gdnoil ain't that the Glo-up when you can't relate goo.gl/bmtxY2

Jennifer Smith @_jsmith1993 · Jul 2
@gap_14 Así la semana 4 del #GA1 de la jornada de #LigaMayor de #Mexico goo.gl/QWbs9t

Jennifer Smith @_jsmith1993 · Jul 2
@Flickie93 Now live with it! :D goo.gl/xNhrGM

Jennifer Smith @_jsmith1993 · Jul 2
@fiieeya Dah plan sama sama , bila jadi macam ni biasanya dipanggil talam berm... goo.gl/uhvyCv

Jennifer Smith @_jsmith1993 · Jul 2

Man vs. Machine 2 Hour Bake Off

User / Metric	Person	SNAP_R
Total Targets	~200	819
Tweets/minute	1.67	6.85
Click-throughs	49	275
Observations	• Copy/Pasting messages to different hashtags	• Arbitrarily scalable with the number of machines

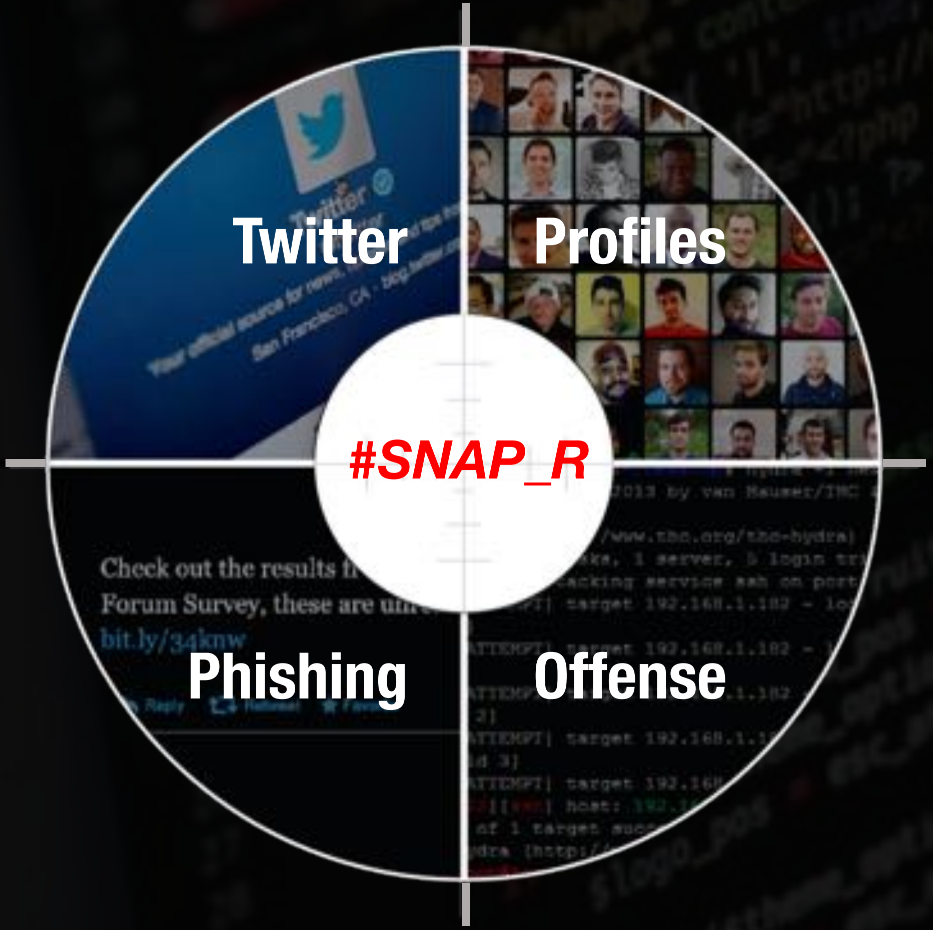
DEMO of #SNAP_R

Wrap Up

Weaponizing Data Science for Social Engineering:

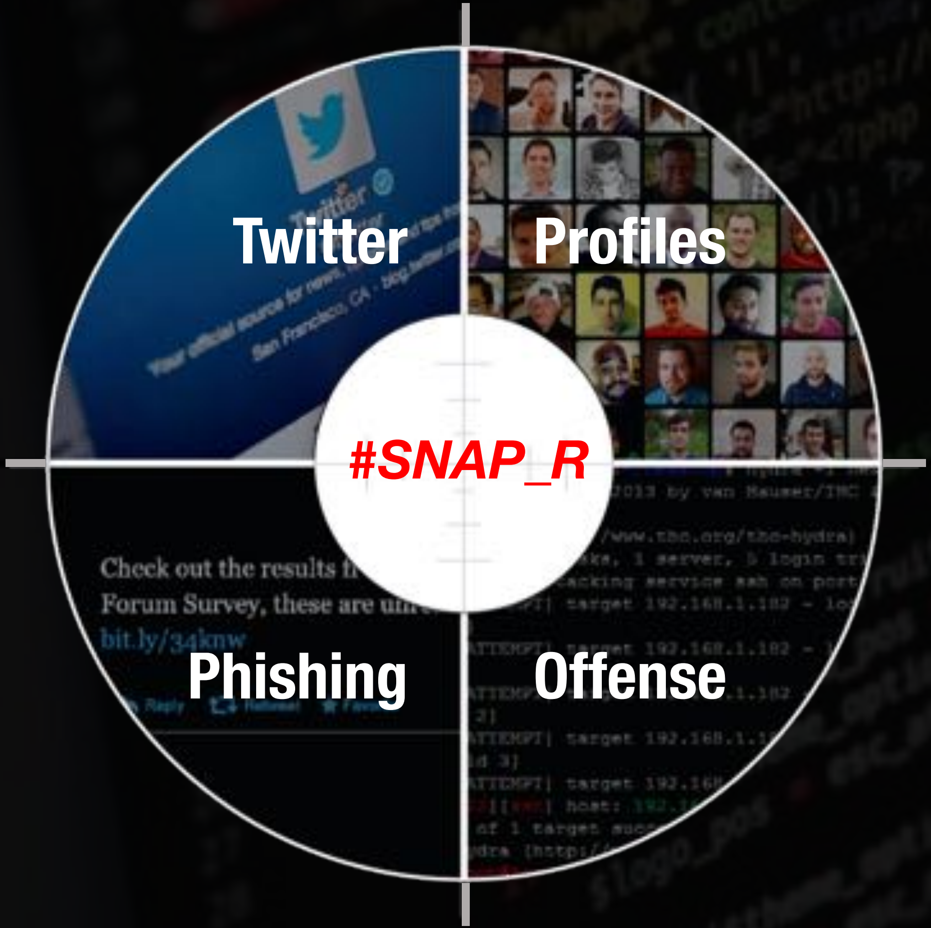
Automated E2E Spear Phishing on Twitter

Potential Use Cases



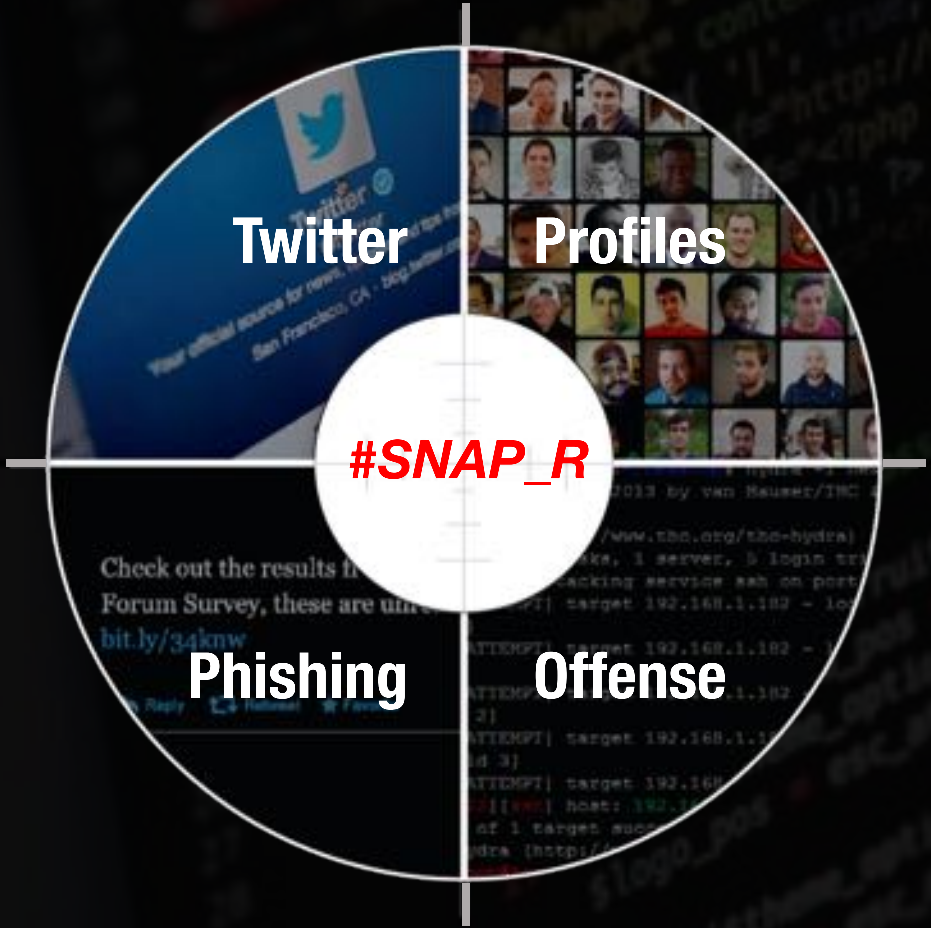
- Social media security awareness
- Social media security education
- Automated internal pentesting
- Social engagement
- Staff Recruiting

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?

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

We'll also be at the ZEROFOX® *booth immediately after the presentation!*