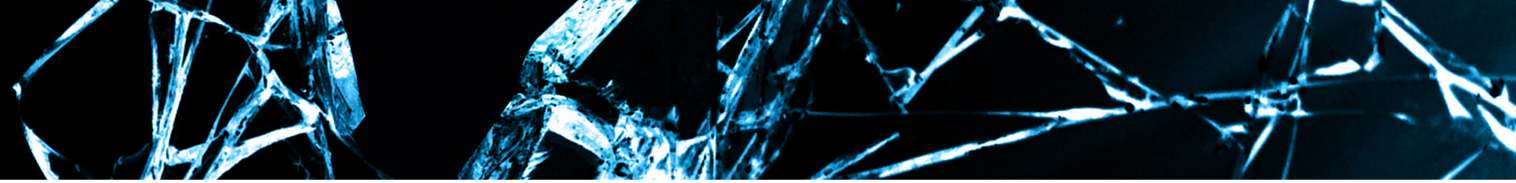
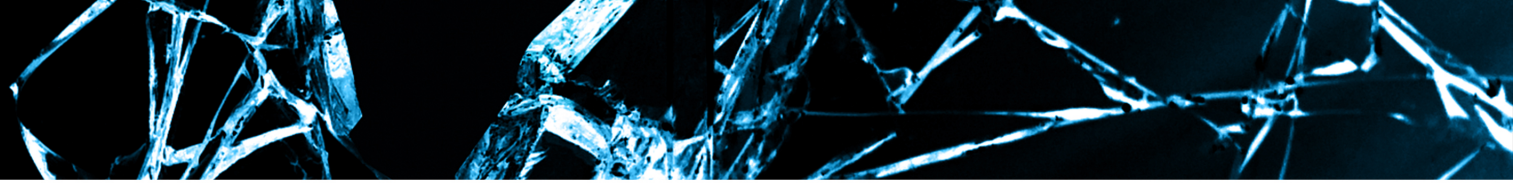

black hat®
USA 2016
JULY 30 - AUGUST 4, 2016 / MANDALAY BAY / LAS VEGAS



# CALL ME: GATHERING THREAT INTELLIGENCE ON TELEPHONY SCAMS TO DETECT FRAUD

Aude Marzuoli, Ph.D.

Pindrop

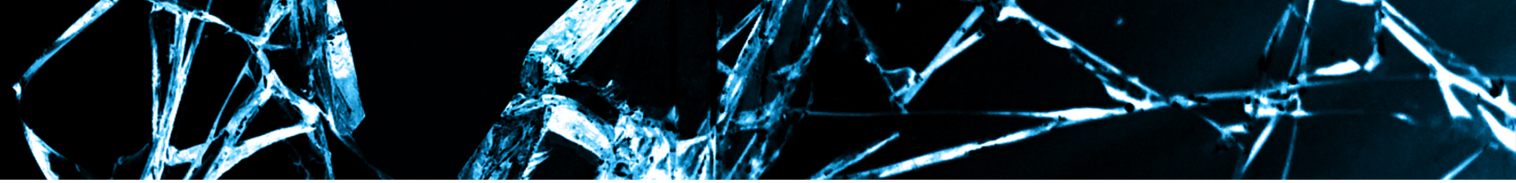


# pindrop

- Pioneer in voice fraud prevention and authentication helping companies eliminate phone fraud.
- Patented Phoneprinting™ technology can identify, locate and authenticate phone devices uniquely just from the call audio.
- Selected by the world's largest banks, insurers, brokerages and retailers, detecting over 80% of fraud, even for attackers never seen before.
- Pindrop was founded in 2011 and to date has raised $122 million.

# Outline

- Introduction to our Honeypot, a.k.a. Phoneypot
- Tracking spam/scam campaigns
- Gathering Intelligence on Telephony abuse
- Conclusion



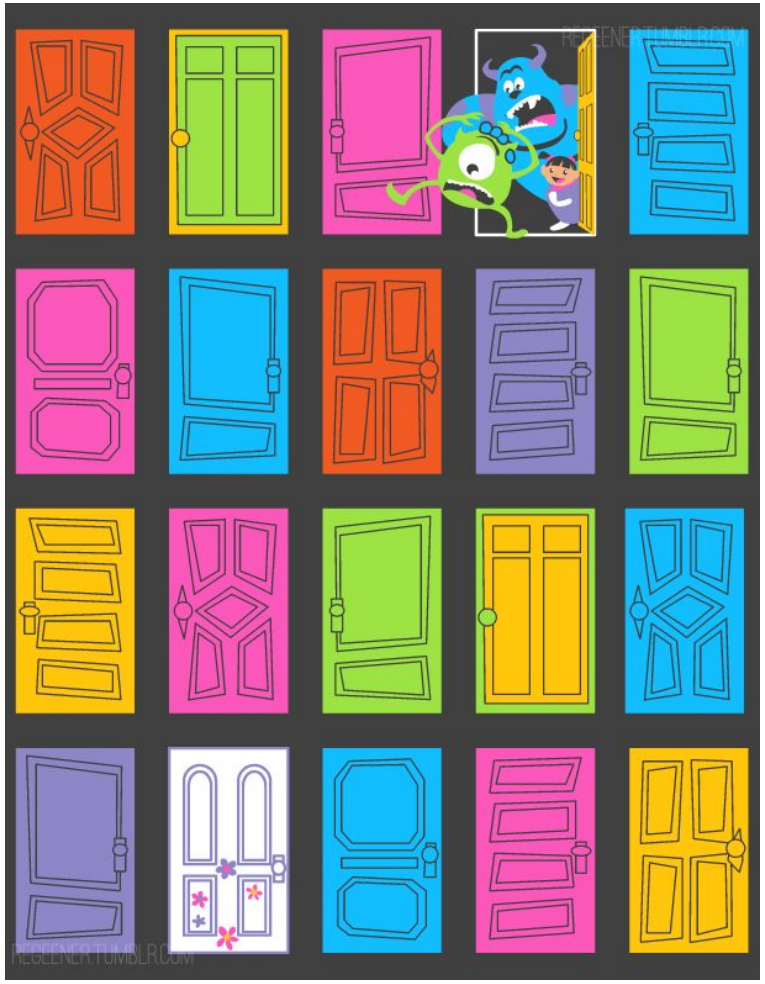
# Introduction to Phoneypot

# Who's REALLY calling you?


# But now scammers are calling you direct


DATA

Customer Names
Financial Institutions,
Account Numbers,
SSN, Account
balance, PIN,
Password, DOB




CNN Money U.S. + Business Markets Tech Media Personal Finance Sma

# Google strikes back at robocalls

ConsumersUnion

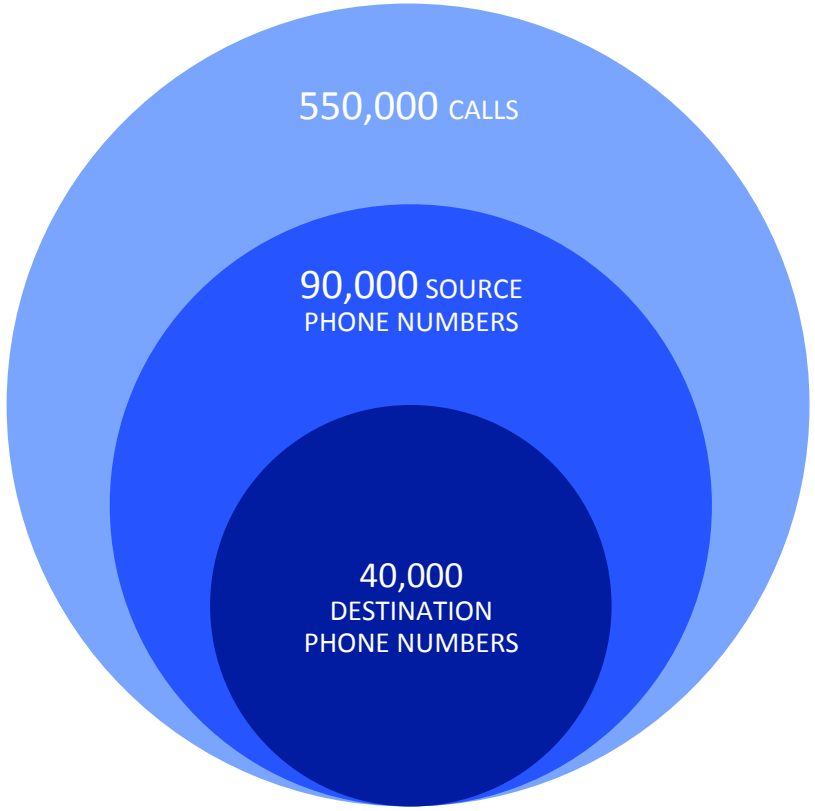
**End Robocalls Activists File Over 15,000 Comments with the FCC: No Robocalls to my Cell!**

# Why research telephony fraud?

- Understand the telephony fraud/scam ecosystem
- Gather threat intelligence through measurements and experiments
- Design defense strategies and deploy them

# Phoneypot

## IN ONE MONTH




## LIMITATIONS

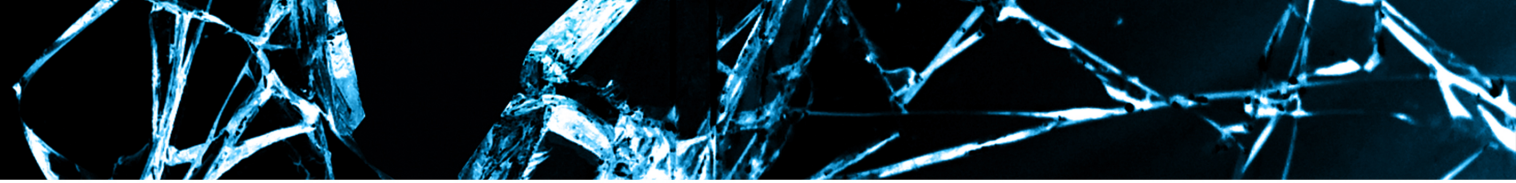
- Source phone numbers spoofing
- Most phone numbers call once or twice
- Labels Needed

During the first half of 2016, we recorded 100,000 calls to our Honeypot.

# PHONEYPOT VS ONLINE COMMENTS

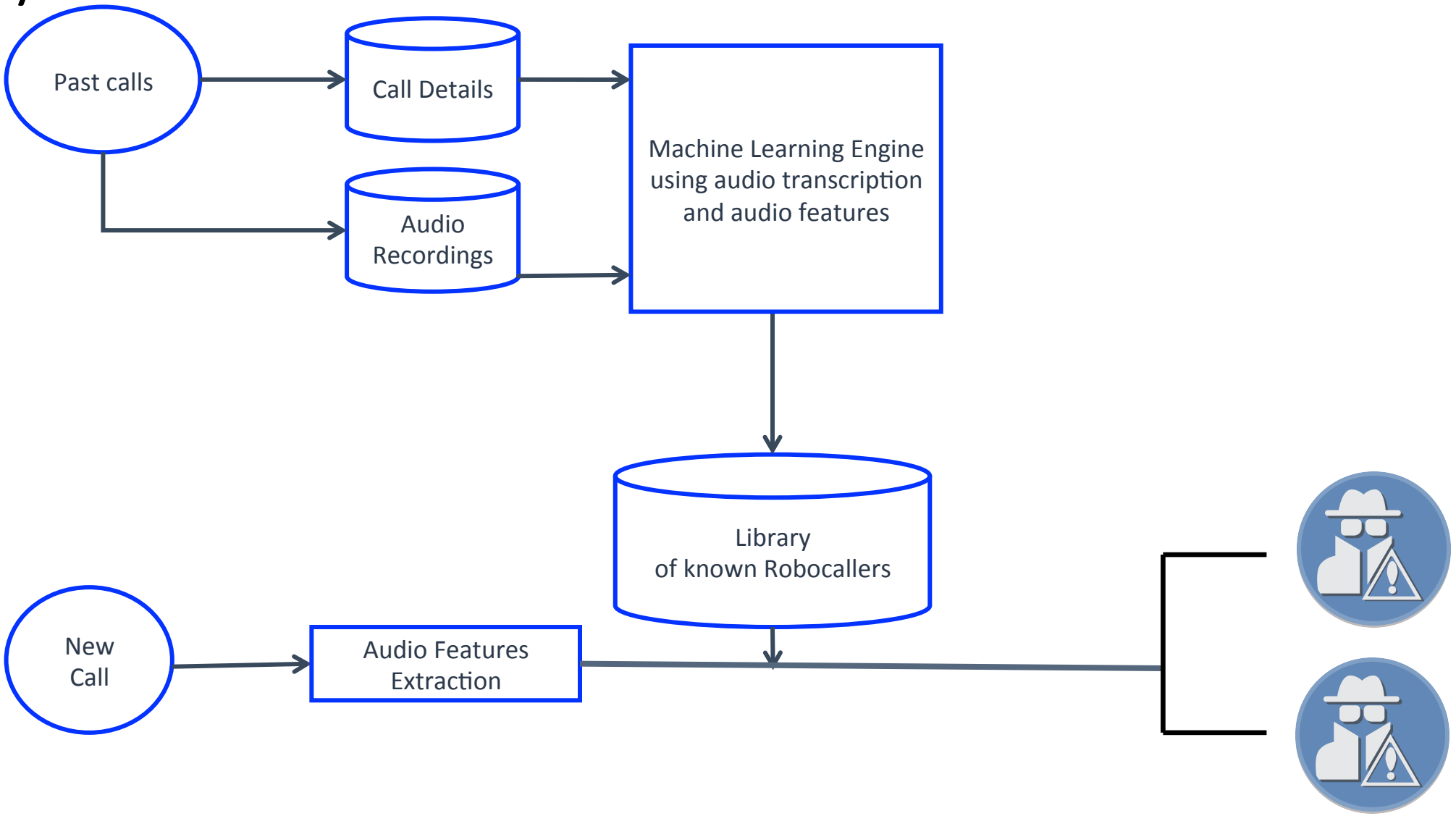
| Data set | # of calls | # of source phone numbers | Max. # of calls or comments per source phone number |
|---|---|---|---|
| Phoneypot | 8,000,000 | 880,000 | 21,329 |
| Online Comments | 660,000 | 74,000 | 2,156 |

- Data set collected for year 2015.
- 66% of online comments are complaining about only 2% of sources in the Phoneypot.



# Tracking spam/scam campaigns

# System Overview



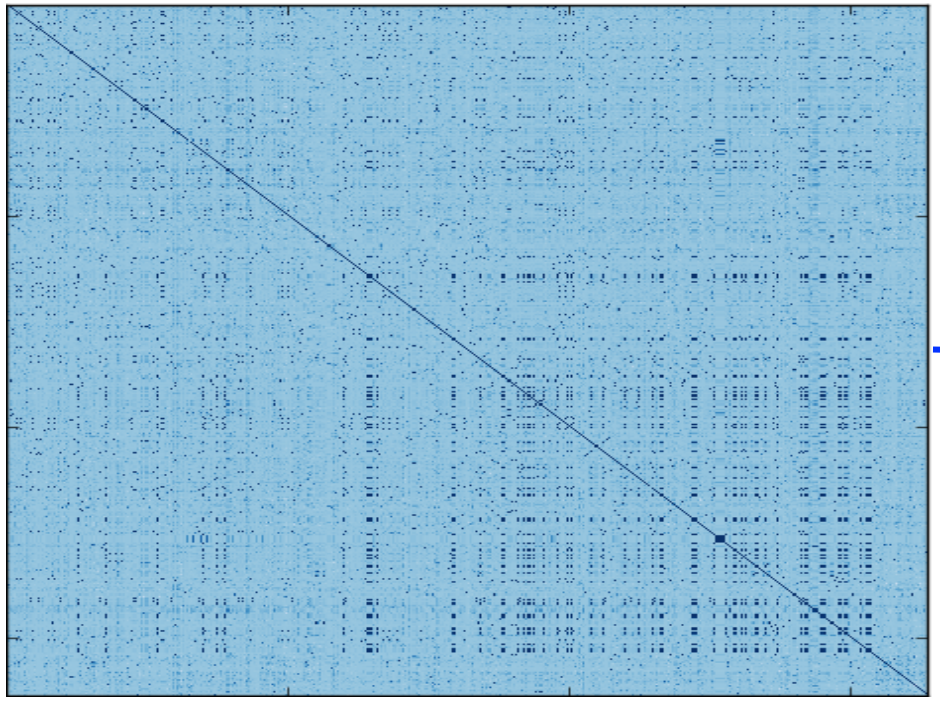
# Keywords are too noisy

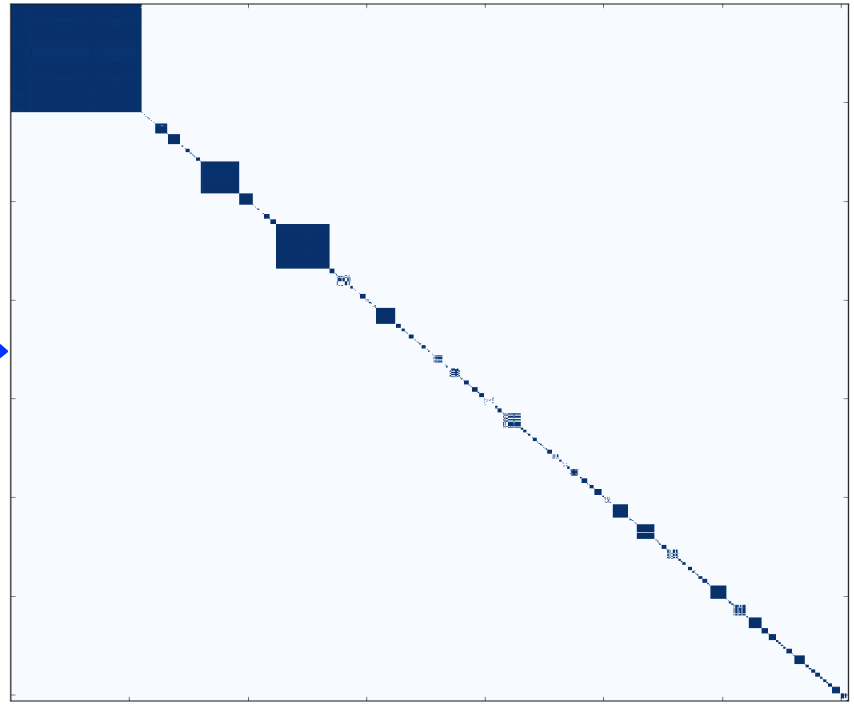
**Transcript to compare:**

- "**press one** to say **fifty per cent** on your **electric bill** with n**o money out of pocket** using in **new jersey** sober **company** with **hundreds of satisfied customers** pressed to and **continue paying one of the highest electric utility rates** "

- **93%** similarity to:
  - "oh **press one** to say **fifty percent** on your **electric bill** with **no money out of pocket** to using a **new jersey**'s older **company** with **hundreds of satisfied customers** pressed to in **continue paying one of the highest electric utility rates**
- **63%** similarity to:
  - "hi we're just doing the quick survey where you where the **new jersey** residence or now entitled to the rebate on their **electric bills** twice a year and a twenty five dollar visa gift card press one"
- **39%** similarity to:
  - "do you have at least ten thousand dollars in credit card debt and medical **bills** would you like to hear about of program that can reduce your bounds or than effectively eliminate your interest "

# Clustering of transcripts similarity matrix

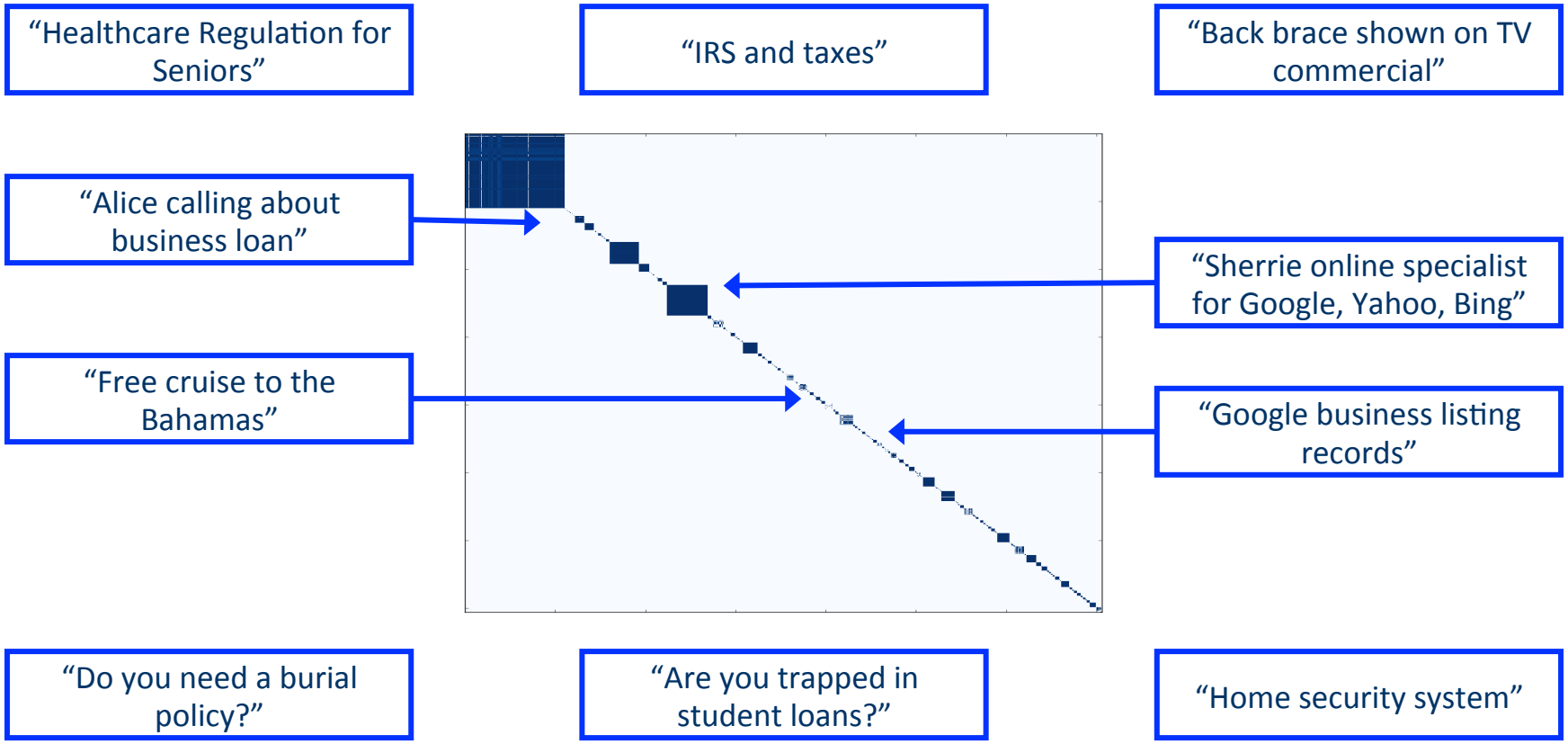




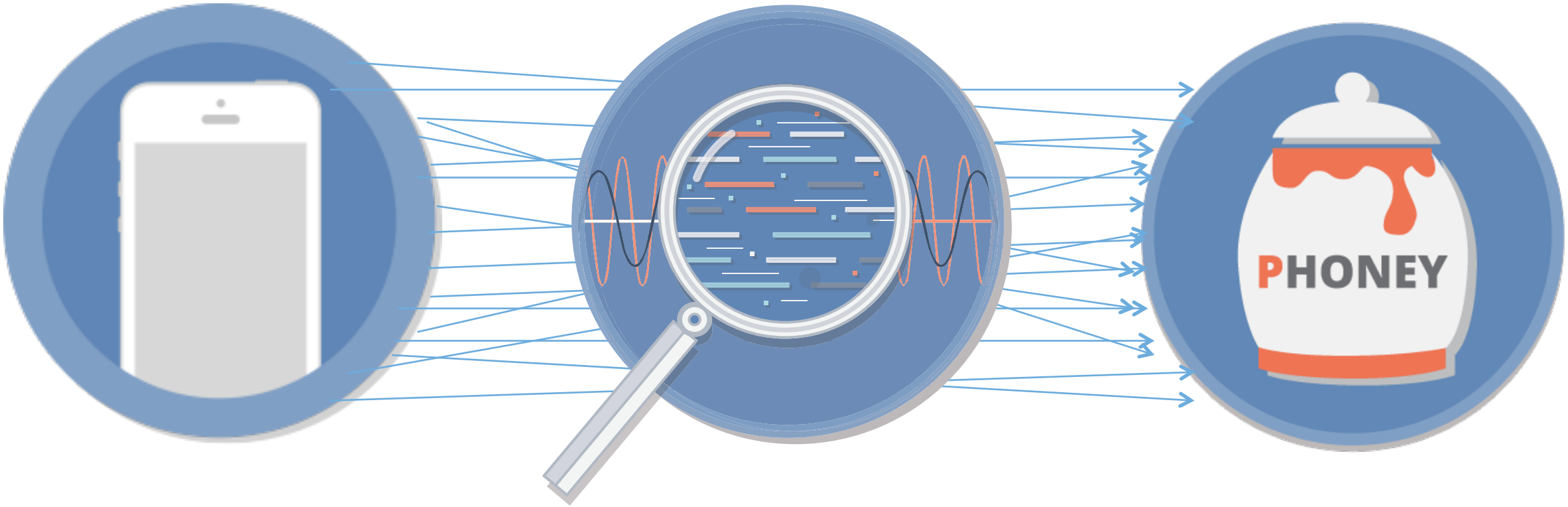
Raw similarity matrix between all pairs of transcripts
Dark dot if transcript i is similar to transcript j

Spectral clustering Results

# Cluster extraction



# Cluster Phoneprinting



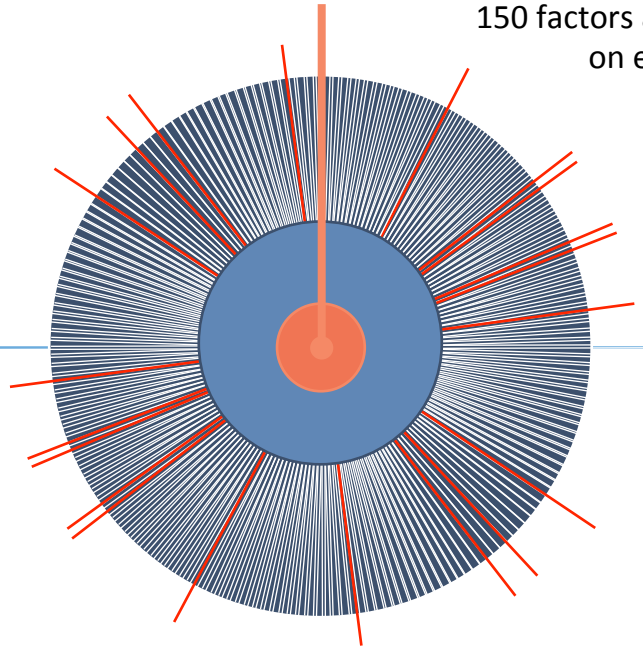

Source phone numbers whose calls correspond to identical audio recordings

Extract **150** audio features for all calls in the same cluster

Phoneypot destination phone numbers

# Phoneprinting

150 factors analyzed on every call



Initial call

Call Type

Geo-Location

Unique Phone

**Spectrum**
Quantization, Frequency filters, Codec artifacts

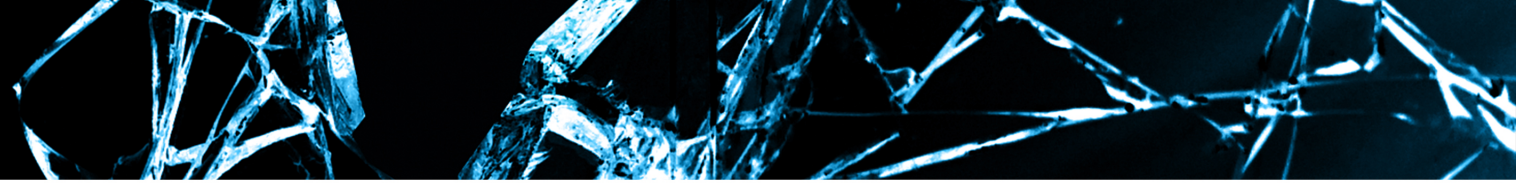
**Noise**
Clarity, Correlation, Signal-to-noise ratio

**Loss**
Packet loss, Robotization, Dropped frames

# Phoneprinting distinct bad actors

- Feed the classifier:
  - Positive class: audio features from all calls in the same cluster
  - Negative class: audio features from any other call
- Tested across more than 50 clusters:
  - Average performance: 85.5% True Positive Rate at 0.25% False Positive Rate
- Good performance even with <2 calls for each source phone number.

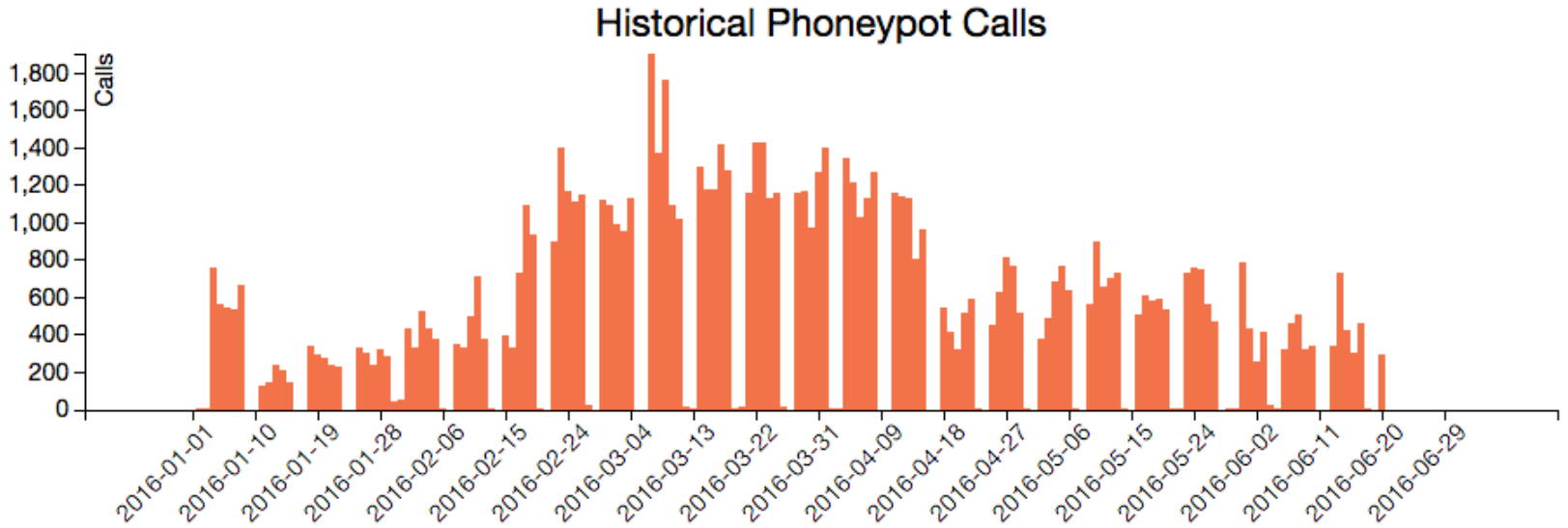
- Good phoneprint performance on a cluster implies that all calls from this cluster come from the same physical telephony infrastructure.

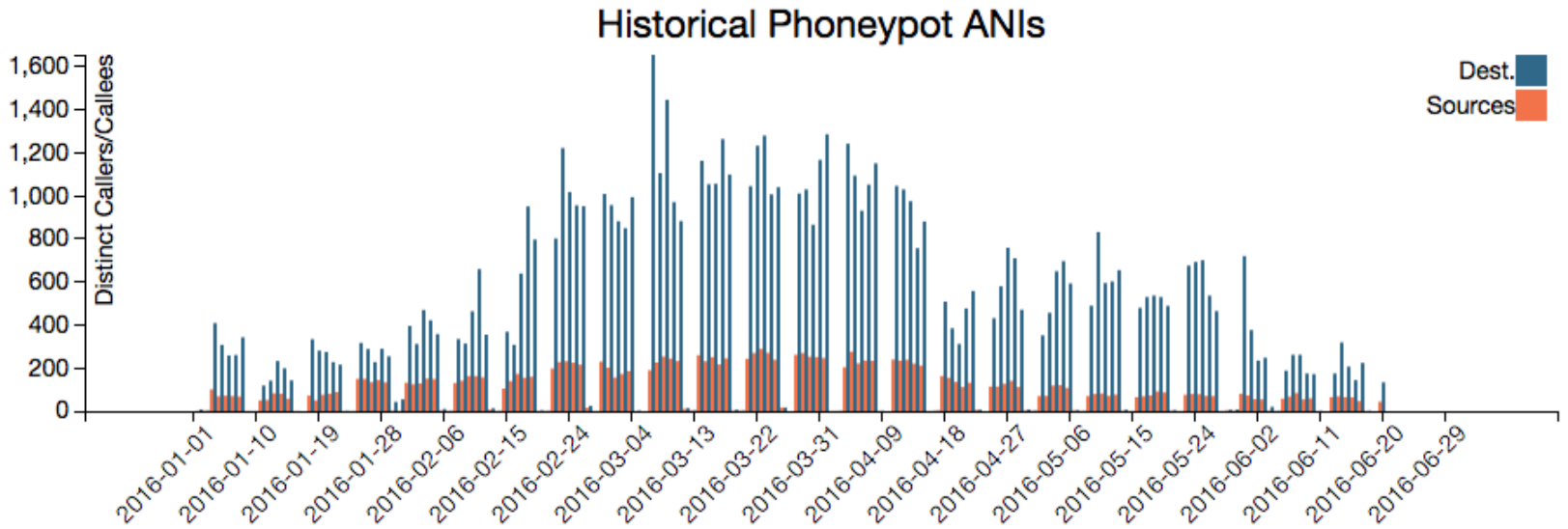


# Gathering Intelligence on Telephony abuse

# Google-related scams

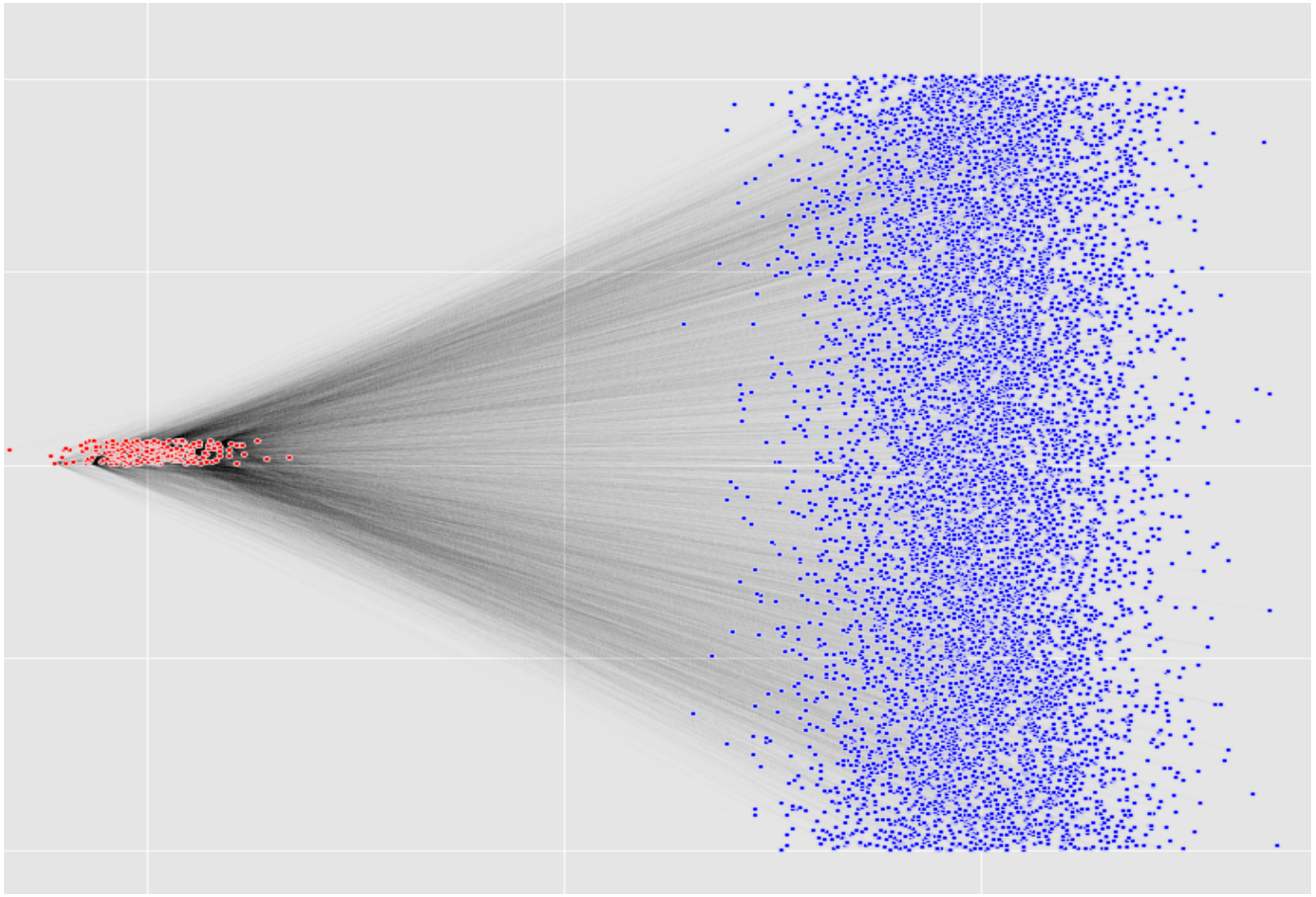
- One of the most complex scams
- More than 10 different recordings targeting Google users, from Google maps to Google plus.
- Asking you to pay money for front page placement on Google search results.
- Asking you to “verify” personal information.




* two of our backers are Google capital and GV

# Sharon, your Google specialist



253 Source phone numbers

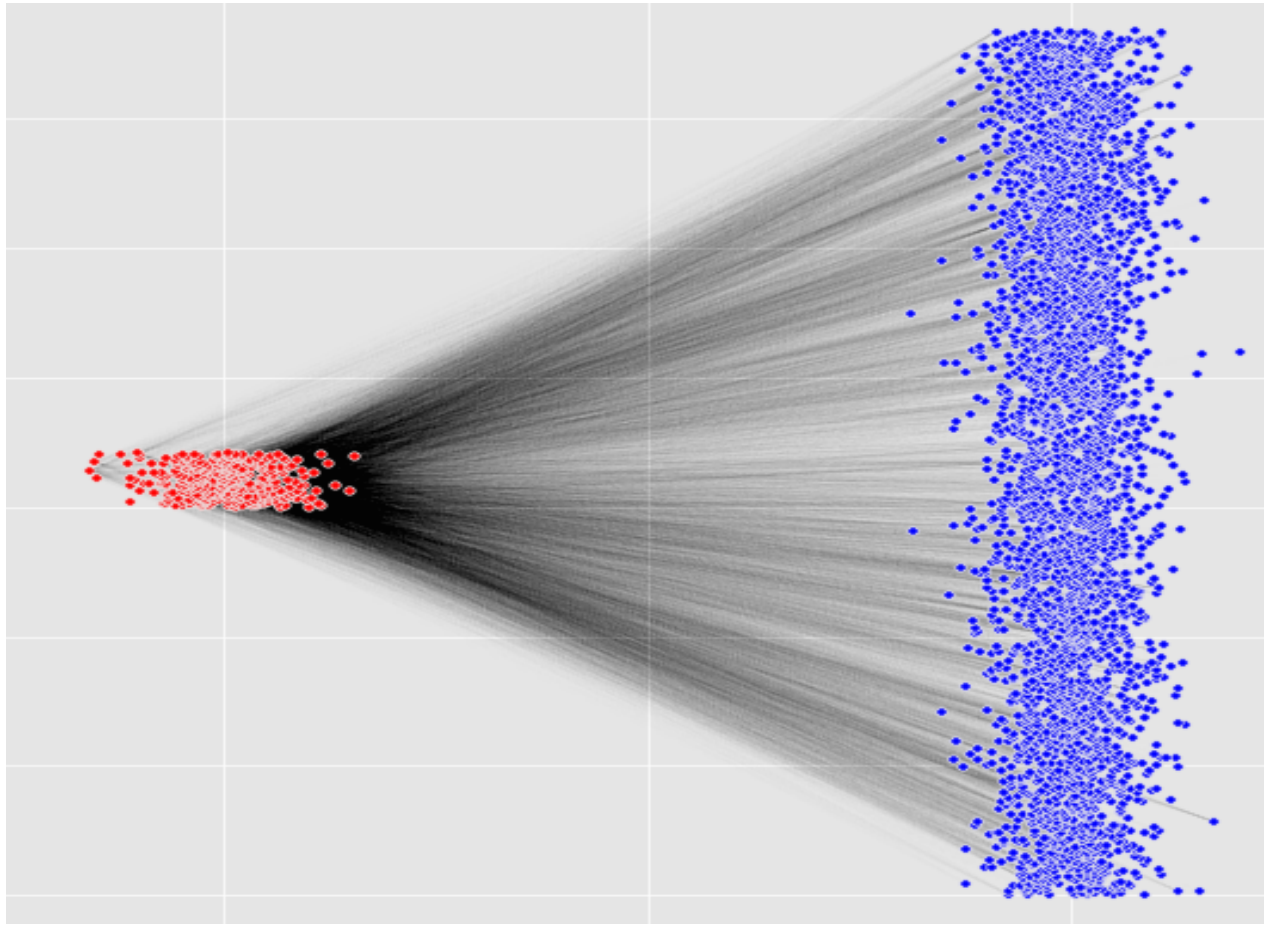
27,928 calls from January to July 2016

17,160 destination phone numbers from Phoneypot

- "Hi this is Sharon your local Google specialist. We have a front page position available for a business like yours and can guarantee front page placement with unlimited clicks 24 hours a day at a flat rate on Google. Press one key right now to see if you qualify and are interested in receiving calls from companies who are locally searching for your type of business. Please press one now or press the two key to be removed. Thank you."

- Phoneprint performance: 93.4% TPR at 0.4% FPR

# Google Business listing



410 Source phone numbers

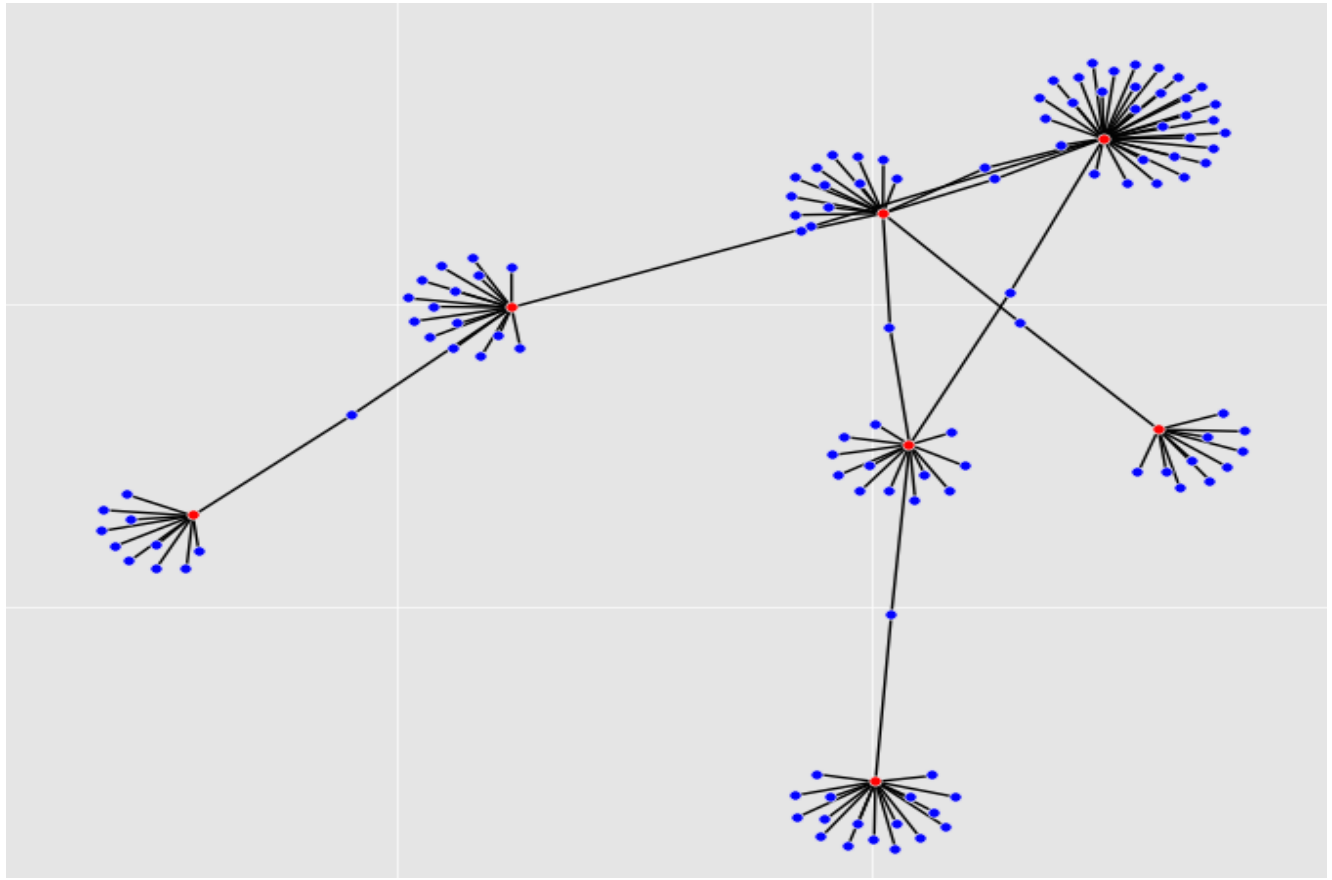
34,048 calls from January to July 2016

26,841 destination phone numbers from Phoneypot

- “This is an important call regarding your google business listing. We need to verify your contact information. Press one now to verify your google business listing. Press seven to remove yourself from the verification system. “

- Phoneprint performance: 92.7% TPR at 0.2% FPR

# Debt Collectors



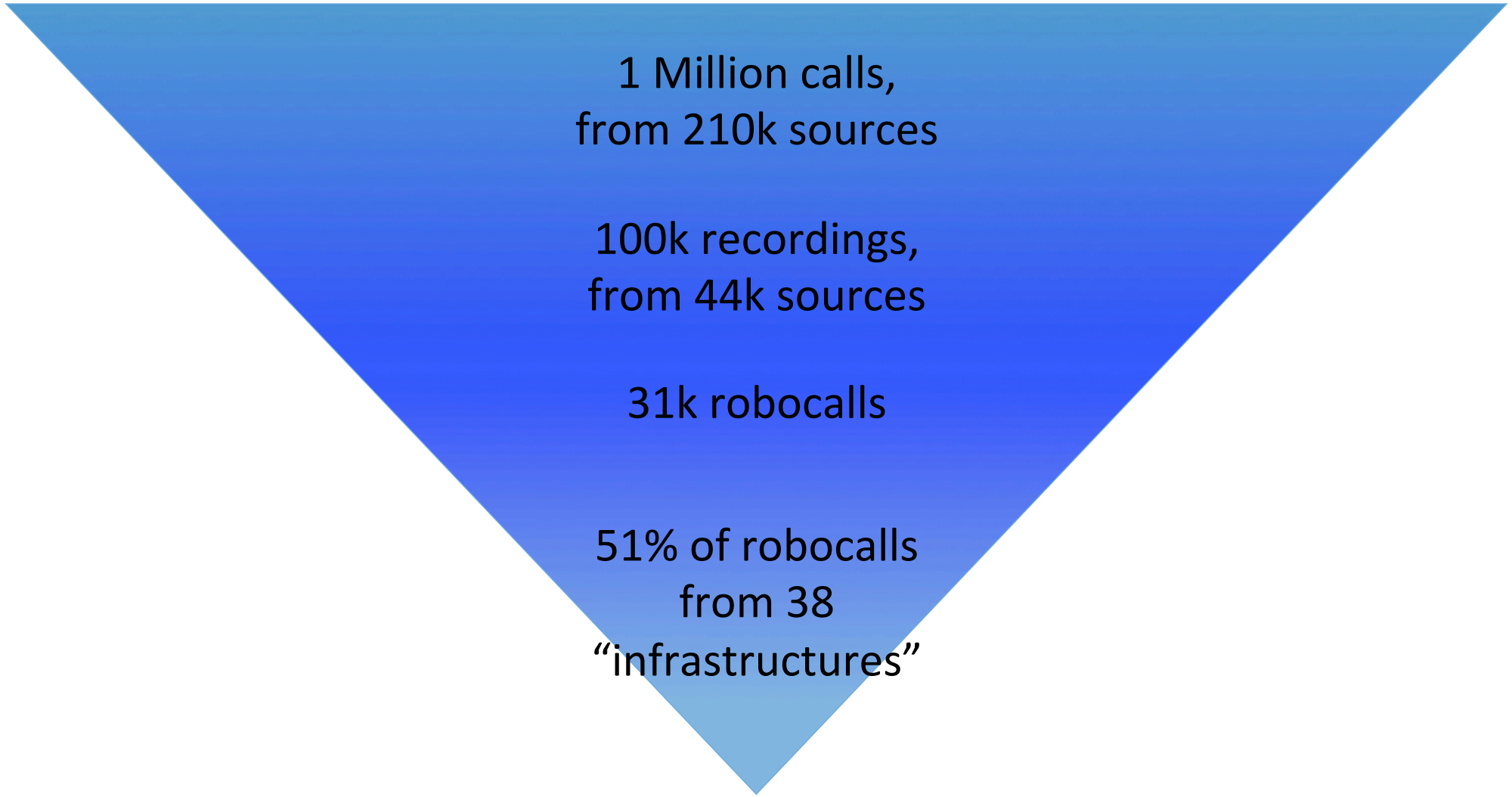
- “This call is from XXXX outsourcing, a debt collector. Please return my call at 866-XXX-XXXX. “
- Targeted harassment tactics

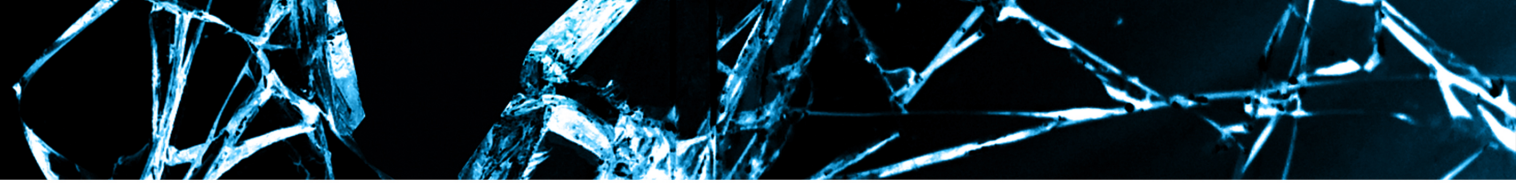
7 Source phone numbers

1,301 calls from January to July 2016

92 destination phone numbers from Phoneypot

# Experiment Overview





# CONCLUSION

# Conclusion

- We recorded 100,000 calls over several months in 2016.
- Using machine learning on call patterns, semantics and audio features, we can uniquely identify one bad actor hiding behind several source phone numbers, whether these numbers are legitimate or spoofed.
- Our results show that 51% of the robocalls recorded are placed by 38 distinct telephony infrastructures which can be uniquely identified with more than 85% TPR on average.

# Acknowledgements & Questions

Data Science team at Pindrop:

Hassan Kingravi, David Dewey, Aaron Dallas, Telvis Calhoun, Terry Nelms